



# CONFERENCE PROCEEDINGS



UGC Sponsored  
National Conference on

## Recent Advancements in Science and Technology

VOLUME IV : Electronics, Computer Science & Application



:: Organized by ::

Vidya Bharati Shaikshanik Mandal's

### Vidya Bharati Mahavidyalaya, Amravati

Re-accredited with Grade 'A' by the NAAC (CGPA 3.23 - Third Cycle)  
College with Potential for Excellence (CPE Status Thrice by the UGC)  
Star College Status by DBT, New Delhi, Mentor College under Paramarsh by UGC  
Identified as Lead College by S.G.B. Amravati University, Amravati

:: In Collaboration with ::

### S.S.S.K.R. Innani Mahavidyalaya

Karanja (Lad), Dist. Washim 444105 (M.S.), India  
NAAC Re-Accredited 'A+' Grade (CGPA 3.28)  
C.P.E. Status Awarded by UGC, New Delhi  
Affiliated to Sant Gadge Baba Amravati University, Amravati

ISBN : 978-81-19931-25-5



# CONFERENCE PROCEEDINGS



UGC Sponsored  
National Conference on

## Recent Advancements in Science and Technology

VOLUME IV : Electronics, Computer Science & Application



:: Organized by ::

Vidya Bharati Shaikshanik Mandal's

### Vidya Bharati Mahavidyalaya, Amravati

Re-accredited with Grade 'A' by the NAAC (CGPA 3.23 - Third Cycle)  
College with Potential for Excellence (CPE Status Thrice by the UGC)  
Star College Status by DBT, New Delhi, Mentor College under Paramarsh by UGC  
Identified as Lead College by S.G.B. Amravati University, Amravati

:: In Collaboration with ::

### S.S.S.K.R. Innani Mahavidyalaya

Karanja (Lad), Dist. Washim 444105 (M.S.), India  
NAAC Re-Accredited 'A+' Grade (CGPA 3.28)  
C.P.E. Status Awarded by UGC, New Delhi  
Affiliated to Sant Gadge Baba Amravati University, Amravati

SAI JYOTI PUBLICATION, NAGPUR

UGC Sponsored  
National Conference on

---

---

**Recent Advancements in Science and Technology**

---

---

**VOLUME IV : Electronics, Computer Science & Application**

---

---

Chief Editor

***Dr. Pradnya S. Yenkar***

Principal, Vidya Bharati Mahavidyalaya, Amravati

**ISBN : 978-81-19931-25-5**

Date : 10<sup>th</sup> Feb., 2024

Publisher :

**Sai Jyoti Publication**

Itwari, Nagpur

E-mail : [sjp10ng@gmail.com](mailto:sjp10ng@gmail.com)

Type Setting & Printing

**LASER POINT,**

Gadge Nagar, Amravati

No part of this book shall be reproduced, stored in retrieval system, or translated in any form or by any means, electronic, mechanical, photocopying and/or otherwise without the prior written permission of the publishers.

• **Disclaimer :**

- The publishers, The Principal, Vidyabharti Mahavidyalaya, Amravati and the organizers are not responsible for the content of any of the research paper/article published in this conference proceedings.
- The authors/Co-authors are responsible for the plagiarism
- For any conflict of interests, the Author/Co-Authors are responsible.
- For any query/issue/ethical aspects contact the corresponding Author.

# CONFERENCE PROCEEDINGS

## UGC Sponsored National Conference on Recent Advancements in Science & Technology

Date : 10<sup>th</sup> Feb, 2024

**:: Organized by ::**

Vidya Bharati Shaikshanik Mandal's

**Vidya Bharati Mahavidyalaya, Amravati**

**:: In Collaboration with ::**

**S.S.S.K.R. Innani Mahavidyalaya**

Karanja (Lad), Dist. Washim 444105 (M.S.), India

### Editorial Board

#### VOLUME IV : Electronics, Computer Science & Application

Chief Editor
<p><b>Dr. Pradnya S. Yenkar</b> Principal, Vidya Bharati Mahavidyalaya, Amravati</p>
Editors

Electronics	Computer Science and Application
<p><b>Dr. R. J. Gajbe</b>, Associate Editor <b>Ms. K. D. Bhanang</b>, Member <b>Mr. Anup Yenkar</b>, Member <b>Mr. N. R. Bundile</b>, Member</p>	<p><b>Prof. Ather Iqbal</b>, Associate Editor <b>Prof. V. N. Mohod</b>, Member <b>Dr. S. R. Thakare</b>, Member <b>Dr. Shilpa B. Sarvaiya</b>, Member <b>Prof. Deven M. Kene</b>, Member <b>Prof. Shital M. Mohod</b>, Member <b>Prof. Dipika S. Harode</b>, Member <b>Prof. P. M. Ingle</b>, Member</p>



# CONTENTS

## Electronic Technology

S.No.	Title	Author/s	Page No.
1	Integrated Fuzzy Logic Controller for Temperature Control	P. A. Saudagar P. V. Tekade S. H. Bagade	1-4
2	Comprehensive Study of E-Waste Hazards and Its Management	Prof. Dr. Gajanan S. Wajire	5-7
3	Iris Detection by using Artificial Intelligence	Dr.R. D. Chaudhari	8-13
4	Design of Node MCU based Data Logger System	R.K.Parate K.M.Dhole S.J.Sharma	14-17
5	Exploring Innovations in Sensor Technologies for Enhanced Healthcare Monitoring Systems: A Comprehensive Analysis	K. Y. Rokde S. S.Shende	18-23
6	An IoT Application- Wireless Water Tank Meter Using Ultrasonic Sensor	Miss. Teesha Das Miss. Aastha Orkey	24-28
7	Artificial Intelligence in Agriculture: An Analytical Study	R. J. Gajbe C. R. Chaudhari A. R. Yenkar N. R. Bundile K. D. Bhanang	29-33
8	A Study on Covid -19 Classification From X-Ray Images With Machine Learning	Siddharth K. Ganvir Dr.G. K. Reddy	34-37
9	AGNIRAKSHAK: Raspberry Pi based fully Robust fire Sensing & Protection System	Siddharth K. Ganvir Dr.S.P.Deshpande Dr.G. K. Reddy	38-41
10	Ladder Logic Diagram – A PLC Programming Method for Automation	Mr. A. G. Kshirsagar Dr. G. D. Agrahari Dr. D. S. Dhote Mr. M. C. Naidu	42-46
11	Third Eye : Arduino Based Aid for Blind	Mr. Nilesh R. Bundile Mr. Chandrakant R. Chaudhari Dr. Rajani J. Gajbe	47-51
12	Wireless E-Notice Board using gsm & at mega 328	Anup R. Yenkar Dr. Rajani J. Gajbe Chandrakant R. Chaudhari Nilesh R. Bundile Kirti D. Bhanang	52-58

---

13	Water Quality Monitoring Using Atmega 328PU	Ms. Bhagyashri Bhuyar Mr. Chandrakant R. Chaudhari	59-66
14	Utilization and Generation of Hydropower for Welfare in Agricultural Sector	Mr. Chandrakant R. Chaudhari Dr. Giridhar K. Reddy	67-71

\*\*\*\*\*

---

# CONTENTS

## Computer Science & Application

S.No.	Title	Author/s	Page No.
1	A Study on the Fog- Edge-Cloud Computing based IoT (FECIoT): Architecture, Security, and Privacy Issues	Prof. Ather Iqbal Dr. C. H. Sawarkar Dr. Shilpa S. Sarvaiya	1-9
2	IoT Node Security Attacks on Device Layer: Attacks Detection Countermeasures and Solutions	Dr. Shilpa B. Sarvaiya Dr. D. N. Satange Dr. A. A. Tayade	10-15
3	Various Approaches for Content Extraction from Web Pages based on Factors	Deven M. Kene Ather Iqbal	16-20
4	Overview and Classification of Social Security Attacks using Online Social Networking for Rumour Blocking	Mrs. Shital M. Mohod Prof. Ather Iqbal	21-27
5	Big Data: Security and Security Challenges	Miss. Dipika S. Harode	28-33
6	Challenges in Devanagari Script-based CAPTCHA: A Comprehensive Analysis	Anita B. Dube	34-39
7	Stress Detection using Machine Learning Techniques	Mrs. M. M. Mohod Dr. P. M. Jawandiyia	40-43
8	IoT Based- Soil Salinity Mapping and Smart Crop Recommendation System	Dr. Avinash B. Kadam Miss. Shubhangi D. Falke	44-50
9	Big Data Analytics In Health Care: A Reviewpaper	Prof. Rana Afreen Sheikh Prof. S. K. Totade	51-59
10	Study of Software Vulnerabilities Detection Tools and Techniques	Mr. Y. V. Hushare Dr. Sudhir B. Jagtap Dr. U. S. Junghare	60-63
11	Trends in Recommendation Engine: A Systematic Review	Sachin N. Joshi Vivek A. Manwar Dr. A. B. Manwar	64-68
12	Machine Learning Tools for Data Science: A Review	Sagar A. Durge Dr. Kishor M. Dhole	69-74
13	A Python Approach For Information Retrieval Using Vector Space Model (VSM)	Narendra. M. Jathe	75-82
14	Model-Driven Design of Graph Databases	Bhushan Jalamkar Dr. S. R. Thakare	83-87
15	Artificial Intelligence for Genetic Mutation Detection and its applications	Bobade A D Bhonde M M Ingle S G	88-93



16	A Research on Cloud Computing	Miss Bhavana K. Bhagat	94-97
17	Impact of Hadoop Technology in Cloud Computing	Prof. Sameena A.Kazi	98-104
18	Internet of Things: Applications and Security Challenges	Miss. Ankita Suresh Chambatkar Miss. Nikita Sakharam Dhande Prof. S. M Mohod	105-109
19	Introduction to Artificial Intelligence & Its Applications	Mr. V. N. Mohod Mr. A. S. Deshmukh	110-115
20	Study of Wireless Communication Technologies with IoT	Ms. Vaishnavi T. Chore Prof. V.N.Mohod	116-120
21	Role of Nanotechnology and Artificial Intelligence in aroma	Ms. Mayuri A. Deshmukh	121-130
22	Using UML in Software Requirement Analysis Case Study of Academic Organization as an Example	Mrs. Anjali R . Shanke (Jadhav)	131-140
23	Artificial Intelligence and it's Applications	Prof. Pratik S. Yawale Prof. Devendra G. Ingale	141-144
24	Cyber Security Awareness on Cyber Attacks	Aparna R.Sapate	145-150
25	IOT Based Vehicle Speed Control by Using Mobile	Prof. Devendra G. Ingale Prof. Pratik S. Yawale	151-154
26	Word Sense Disambiguation for Marathi Language in Cross Language Information Retrieval	Vivek A. Manwar Rita L. Gupta Dr. A. B. Manwar	155-158
27	Machine Learning Applications: A Comprehensive Overview of Techniques and Working Mechanisms	Ms. Vaishnavi S. Karale Dr. Priyanka C. Tikekar	159-163
28	Intricacies of VR Features in graphic designing	Dr. Bhargavi S. Chinchmalatpure Mr.Pratik Vilas Dabherao	164-168
29	Review Paper on Characteristics, Benefits and Challenges in Cloud Computing	Miss. Sidhdee Satish Gurjar Miss. Adika B. Thakare Prof. D. M. Kene	169-173
30	A Review on Big Data Challenges and Hadoop Technology	Kavita Kishor Yadav	174-179
31	A Review on Internet of Things (IoT) Sensors	Mithilesh M. Wasu Dr. Kishor M. Dhole	180-184
32	A Survey on Predictive Analysis of Social Media Data Using Machine Learning Algorithm	Ms. Sheetal M. Yawalkar Prachi Mawale	185-192
33	A Systematic approach for Challenges and Preventive Measures for Machine learning technology	Ms. Ashwini S. Kaware	193-196

34	Artificial Intelligence: Advanced Analysis and Design	Miss.Shrutika Ramesh Dharamkar Miss.Gopika Shyambabu Ghare Miss.Aparna R. Sapate	197-201
35	Disconnected Realities: Unveiling the Challenges and Impact of Network Issues in Mobile Computing	Nitin Babaraoji Vasu	202-205
36	HTML5 in Web Development	Varun Sanjay Shende	206-208
37	Itemized Selection of Appropriate Image Steganography Method	Prof. V. M. Jawade Prof. R. R. Bhale	209-212
38	Impact of Skill enhancement Programme on Developing Multimedia Tools for Innovative Learning Solution	Ravisha R. Ambekar	213-217
39	IoT-based Automated Smart Irrigation System Using Sensors for Farming	Dr. Avinash B. Kadam Mr. Chandrakant R. Patorkar Miss. Shubhangi D. Falke	218-222
40	Data Mining Algorithms Analysis on Weka for Disease Classification	Sushilkumar R. Kalmegh Prerna S. Tayade	223-229
41	Cloud Computing: Types, Security Issues, Benefits	Miss. Rutuja Naresh Turkhade Miss. Nirja Vinod Deshmukh Dr. Shilpa B. Sarvaiya	230-234
42	Pre-Processing techniques Using Weka Tool	Priya Janardhan Deshmukh	235-239
43	Web Data Mining: A Comprehensive Analysis of Types, Tools, and Techniques	Ms. Anjali V. Parasmode Dr. Sonali R. Chavan	240-243
44	Feature Analysis of Fake News: Improving Fake News Detection on Social Media	Pooja. G. Borkar	244-248
45	Role of Internet of Things in Disease of Pate and Remote Monitoring System	Rashmi A.Charjan Dr.Ajit Kumar Dr.G.K.Reddy	249-252
46	Security IoT Device Against Emerging Security Threats: Challenges and Solution Techniques	Madhumita Y. Sugandhi	253-262
47	Using Machine Learning to Enhance Security Measures at the Network Layer of IoT	Ms. Varkha K. Jewani (Ms. Pragati V. Thawani) Dr. Prafulla E. Ajmire Ms. Geeta N. Brijwani	263-268
48	The Role of chatGPT and other Artificial intelligent in the field of Renewable Energy and sustainable Energy	Mr.Anup Satishrao Bele	269-271

49	An In-Depth Exploration of AI in the Digital Age- A Focus on Scams and Frauds in the Introduction Phase	Ms. Rhutika V. Jawarkar Dr. Priyanka C. Tikekar	272-275
50	Biometric Authentication & its Security Purposes	Prof.S.B.Bele Prof. K.P. Raghuvanshi	276-280
51	Importance of Cyber Security and Cyber Security Tips	Mrs. Rekha N. Yeotikar	281-285
52	A Survey on Blockchain Technology Concepts, Applications and Security Issues	Jaykumar Meshram Dr. Dinesh Satange Dr.Swapnil Deshpande	286-290
53	A survey on face detection techniques using deep learning	Prof.Amit B.Rehapade Dr P. E. Ajmire Kiran H. Varma	291-295
54	A Deep Dive Into Extended Reality -Six Sense Integration-Merging Real And Virtual World	Miss. Sidra-tul-munteha mohd Sabir Miss. Afifa gulam ahmad Saudagar Prof. Dipika S.Harode	296-300
55	Cyber Security: Hacking, Child Pornography, Virus Dissmination	Miss. Gayatri Dilip Wange Miss. Diksha Laxman Pandey Miss. Mayuri Arun Deshmukh	301-305
56	Novel Design and Implementation of the Personalized Search Engine in Context with User Keywords Profile and Keywords Optimization Technique	Mr. K.P.Raghuvanshi	306-310
57	How biometric affects Cyber Security	Miss. Vaishnavi Govardhanrao Dhole Miss. Sakshi Sanjay Rathi Prof. Dr.Shilpa B.Sarvaiya	311-314
58	Analyzing the Challenges ofMarathi Textual Data for Sentiment Analysis	Ram B. Ghayalkar Prof. Dr. D. N. Beseekar	315-319
59	Biometric its application and challenges in Internet of Things	Aloksingh M. Thakur Chandrakant R. Mankar Dr. V. M. Patil Dr. D.N. Satange	320-321
60	Automatic Sparse Representation for Classification of Echocardiographically Detected Intracardiac Masses	Dr. A.A. Tayade Prof. P.M. Ingle	322-326
61	EEG Signal Processing for Fetus and Mother Using MATLAB using Image Processing	Dr. N.D. Jambhekar Dr. R.K. Nawasalkar	327-331

\*\*\*\*\*

# **Electronics**



## 1

## Integrated Fuzzy Logic Controller for Temperature Control

P. A. Saudagar<sup>1</sup>, P. V. Tekade<sup>2</sup>, S. H. Bagade<sup>3</sup>,

<sup>123</sup> Bajaj College of Science, Wardha, India

<sup>1</sup> saudagar.pa@gmail.com

<sup>2</sup> pradiptekade@gmail.com

<sup>3</sup> sanjaybagade8@gmail.com

### Abstract

This paper deals with the development of intelligent controller based on Fuzzy Logic and PID to control the temperature. Thirty-three fuzzy rules were used for the control action to be taken by the controller to maintain the temperature at the set point value. This controller is facilitated with the handheld module for the user to set the parameters as per the need and requirement of the crop. The fuzzy logic is more preferred in control applications.

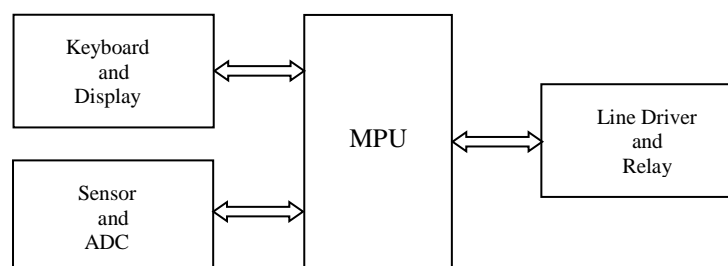
**Keywords** - IFLC, Fuzzy Inference, Fuzzy Logic, PWM

### 1. INTRODUCTION

In many technical processes, an efficient controller for the surrounding environment is vital. From fabrication of Integrated Circuits to the production of chemical solutions, any changes in the ambient parameters can have a drastic effect in the outcome of a process and ultimately on the quality of the product. Temperature is one of the main parameters to control in most of the industries and other fields. In these kinds of industries, some product needs the required temperature to be maintained at highest priority.

In industry, several conventional methods for controlling temperature are available but are less effective as they either based on proportional methods or on-off control. Fuzzy logic has been widely used in industrial controls and domestic electrical equipment [1]. The automatic learning of fuzzy rules is a key technique in fuzzy control. In the present work, a controller was designed which controls the temperature of a plant by continuously monitoring the temperature and comparing it with the set value. An integrated fuzzy logic [5][6] approach was used to decide the output.

### 2. SYSTEM BLOCK DIAGRAM



#### 2.1 HARDWARE

The hardware of Integrated Fuzzy Logic Controller (IFLC) was designed using Atmel's 89C52 microcontroller as an MCU, which initialize the system, reads the sensors, displays the values on LCD and take decisions according to the algorithm. IC LM 35 was used as a temperature sensor to sense the inside temperature of the plant whose outputs were converted to digital using 8-bit Analog-to-D converter IC. An LCD display module was used to display the current values of the temperature and some other messages to have more interactivity. The port 1 pins

P1.0 and P1.1 were used for controlling the heating system. These port lines were further connected to the 10A SSRs through line drivers. The PWM output was generated at the port line to control the heating system.

## 2.2 SOFTWARE

The software for the system was designed in modular form so as to upgrade the required module in future. The software broadly consists of following modules: Initialization Module, Sensor module, Keyboard and display module, I<sup>2</sup>C Bus module, IFLC module and PWM generation module. On manual or power-on- reset, reset, the initialization module loads the variables, stack, and other necessary registers to their default values set by the programmer, initialize the timers and start them. The inside and outside temperatures are sensed by sensor module then converted it to digital form by A/DC and saves them to the assigned address. The keyboard and display routine allows the user to set the temperature inside the plant also displays the settings and other parameters to the LCD. The I<sup>2</sup>C bus is used to save the set values and other parameters as a backup in case of power failure. So, every time user modifies any data it automatically saves in this memory. In this way plant starts its functioning after power resumes. The internal 16-bit timer of microcontroller was used for generation of PWM outputs for the heating system.

## 3. INTEGRATED FUZZY LOGIC CONTROLLER (IFLC)

In the present work, an integrated fuzzy logic controller (IFLC) [2][3][4] was used for maintaining the temperature of the plant to the value set by the user. The IFLC is an integration of a fuzzy logic controller and the PID controller. The block diagram of integrated fuzzy logic controller is shown in Fig. 3.1 An FLC was designed to which two input variables  $e_{temp}$  and  $ce_{temp}$  were given. The  $e_{temp}$  is the error value of temperature, which was computed as-

$$e_{temp} = \text{temperature set point} - \text{current value of temperature}$$

and change in error  $ce_{temp}$  is computed as-

$$ce_{temp} = \text{current } e_{temp} - \text{previous } e_{temp}$$

As the values of  $ce$  and  $e$  are crisp in nature, it needs to be converted to fuzzy values. The triangular membership function was used for fuzzification of  $e_{temp}$  within the universe of discourse with eleven linguistic values, the linguistic values were NVVL (Negative Very Very Large), NVL (Negative Very Large), NL (Negative Large), NM (Negative Medium), NS (Negative Small), Z (Zero), PS (Positive Small), PM (Positive Medium), PL (Positive Large), PVL (Positive Very Large) and PVL (Positive Very Very Large). The universe of discourse for error was  $(-40, +40)^{\circ}\text{C}$ .

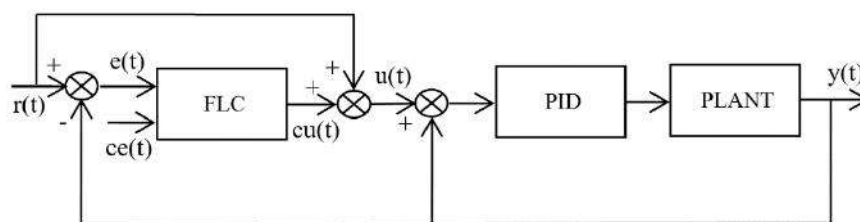


Figure 3.1 Block diagram of IFLC

For  $ce_{temp}$  also triangular membership function was used for converting it to non-crisp values with three linguistic values NEG (Negative), Z (Zero) and POS (Positive).

The important section of fuzzy control was designed from the expert knowledge, experience and previous work in this field. The decision-making stage consists of fuzzy rules to decide what action needs to be taken. The rule base consists of 33 rules, as shown in Table 3.1.

Table 3.1 fuzzy logic rule base

ca ce	e	NVVL	NVL	NL	NM	NS	Z	PS	PM	PL	PVL	PVVL
NEG		VH	VH	H	LH	M	VL	M	LH	H	VH	VH
Z		VH	VH	VH	H	LH	VL	LH	H	VH	VH	VH
POS		VL	VL	VL	VL	VL	VL	VL	VL	VL	VL	VL

The selection of either cooling or heating system was based on the error value. If error is positive the heating system will be selected otherwise cooling system was selected and control action would be applicable to selected system.

If-Then rules are used to take control action. For example, IF  $e_{temp}$  is NM AND  $ce_{temp}$  is NEG then control action is LH. This process is known as inference. The inference process relates the fuzzy state variables  $e_{temp}$  and  $ce_{temp}$  to the fuzzy control action  $ca_{temp}$ . The triangular membership function was used for fuzzification of control action having linguistic values VL (Very Low), L (Low), M (Medium), LH (Little High), H (High) and VH (Very High).

The fuzzy rule-based Mamdani inference [1][5] is used for decision making. These outputs are fuzzy values, which are then converted to crisp value in defuzzification stage. For defuzzification, the fuzzy values obtained by decision-making stage were converted into non-fuzzy or crisp value, the Center of Gravity (COG), also known as Center of Area or Centroid, method was used for this purpose. This method has proved to work well with efficient and accurate result [6].

The crisp value obtained from defuzzification stage for the corresponding values of  $e_{temp}$  and  $ce_{temp}$  were scaled and stored in the internal ROM as look up table and the values of  $e_{temp}$  and  $ce_{temp}$  were used to access these crisp values from the look up table. The value read from the look up table was added to the temperature set point and it was treated as the new set point for PID controller as shown in system block diagram. The current temperature value was subtracted from the new set point value and the difference was considered as error value  $e_n$  for PID controller. The PID controller generates the new control action  $v_n$  implemented using the velocity equation [3][4]-

$$v_n = v_{n-1} + kp(e_n - e_{n-1}) + ki(e_n)T + kd/T(e_n - 2e_{n-1} + e_{n-2})$$

where,

$v_n$ : Control action by PID controller

$v_{n-1}$ : previous control action

$e_n$ : current error value of PID controller

$e_{n-1}$ : previous error value of PID controller

$e_{n-2}$ : previous to previous error value to PID controller

$kp$ : proportional gain

$ki$ : integral gain

$kd$ : differential gain

$T$ : cycle time

For the present controller, the values of  $kp$ ,  $ki$  and  $kd$  were finalized as 2, 1 and 1 respectively. After reset, previous values of error and  $v_n$  were initialized to zero. The value computed by the above equation was scaled within the limits and the scaled value was used for calculation of percentage duty cycle for PWM output. This procedure was repeated to maintain the temperature of plant as per the set point.

#### 4. CONCLUSION

The complete system was tested for its performance. The user interface worked well allowing several parameters to be set as per the need. Once the set point is made, the software acts



---

accordingly and in IFLC generates the PWM outputs necessary for maintaining the set temperature in the plant. The Current temperature values and the set points are continuously displayed on the screen. The IFLC shows improves performance with respect to stability in maintaining the temperature.

## 5. REFERENCES

1. John Yen, Reza Langari, *Fuzzy Logic Intelligence, Control and Information* (Pearson Education, 2003).
2. Ming-Yuan Shieh and Tznu-Hseng S. Li, Integrated Fuzzy Logic Controller Design, *IEEE 1993*, pp279-284.
3. S. S. Patil and P. Bhaskar, Design and Real Time Implementation of Integrated Fuzzy Logic Controller for a High Speed PMDC Motor, *International Journal of Electronic Engineering Research, Vol.1 No.1 (2009)*, pp13-25.
4. S. S. Patil, P. Bhaskar and L. Shrimanth Sudheer, Design and Implementation of An Integrated Fuzzy Logic Controller for a Multi-Input Multi-Output System, *Defence Science Journal, Vol. 61, No. 3 May 2011*, pp219-227.
5. Timothy J. Ross, *Fuzzy Logic with Engineering Applications* (Wiley-India, 2005).
6. Scott S. Lancaster and Mark J. Wierman, Empirical Study on Defuzzification, *IEEE (2003)*, pp121-126.
7. P. A. Saudagar, D.S. Dhote and K. D. Chinchkhede, Design of Fuzzy Logic Controller for Humidity Control in Greenhouse, *International Journal of Engineering Inventions* ISSN: 2278-7461, ISBN: 2319-6491 Volume 1, Issue 11 (December2012) PP: 45-49
8. Saudagar, P. A., Fuzzy and Integrated Fuzzy Logic Controller for Greenhouse Climate Control (Unpublished doctoral thesis). Sant Gadge Baba Amravati University, Amravati, India (2013).

## **Comprehensive Study of E-Waste Hazards and Its Management**

**Prof. Dr. Gajanan S. Wajire**

Department of Electronics, Shri Shivaji College of Arts, Commerce & Science, Akola - 444003 (M.S.)

e-mail : gsw.741@gmail.com

### **Abstract:**

Electronic waste, or e-waste, is a rapidly growing environmental concern as technological advancements lead to increased consumption of electronic devices. This research paper aims to provide a comprehensive overview of the hazards associated with e-waste and the various strategies for its effective management. The paper discusses the environmental, social, and health impacts of improper e-waste disposal and highlights the importance of adopting sustainable practices to mitigate these hazards. Also, it focuses on possible remedies to control e-waste.

Now a day, an importance of the new concept 'Work from home' has become a recent hash tag of the world. As a result, the need of all kinds of electronic devices, gadgets and media is consistently growing, which is insisting more burden of e-waste in the world. Thus, e-waste management plays an important role in the process of sustainable development for human as well as environmental issues.

**Keywords: E-waste hazards, impacts and strategies, legislations, recycling & disposal.**

### **1) Introduction:**

The proliferation of electronic devices in contemporary society has resulted in an alarming surge in electronic waste. E-waste encompasses discarded electronic appliances, gadgets, and equipment, containing hazardous materials that pose significant risks to the environment and human health. This paper addresses the urgent need for a systematic approach to manage e-waste, considering its detrimental effects on ecosystems and public well-being. The concept of e-waste is still new in India, but it is very important and is true need of the hours.

### **2) Hazards and impacts of E-Waste:**

Electronic waste material contents number of toxic and hazardous substances like Tin, Lead, Mercury, Sulphur, Cadmium, Beryllium, Chromium, Americium, plastics and many other oxides. These substances create various environmental problems and damages through emission as well as convention processes. Toxic chemicals from e-waste can pollute the soil, crop and food sources. Also, these are non-biodegradable and causes air & soil contamination as well as pollution. E-waste dumping backyards and nearby places are polluted and causes health hazards to human and animal beings.

Various impacts of e-waste can be listed as below:

- i) Environmental Impact of e-waste results in soil contamination, water pollution and air pollution too.
- ii) Health Impact on human being due to exposure to toxic substances includes respiratory, reduced fertility, slower growth rate and skin ailments.
- iii) Long-term health implications may include liver damage, kidney damage, heart damage, eye and throat irritation.
- iv) Social Impact due to informal recycling and its consequences affects the vulnerable communities in various aspects.

### 3) E-Waste Management Strategies:

E-waste get generated from various resources and sectors of the society. India is 5<sup>th</sup> largest E-waste producer country in the world; after China, USA, Japan & Germany. As per the Indian survey of last year, the sector wise percentage of collected e-waste is pictured in following figure-1.

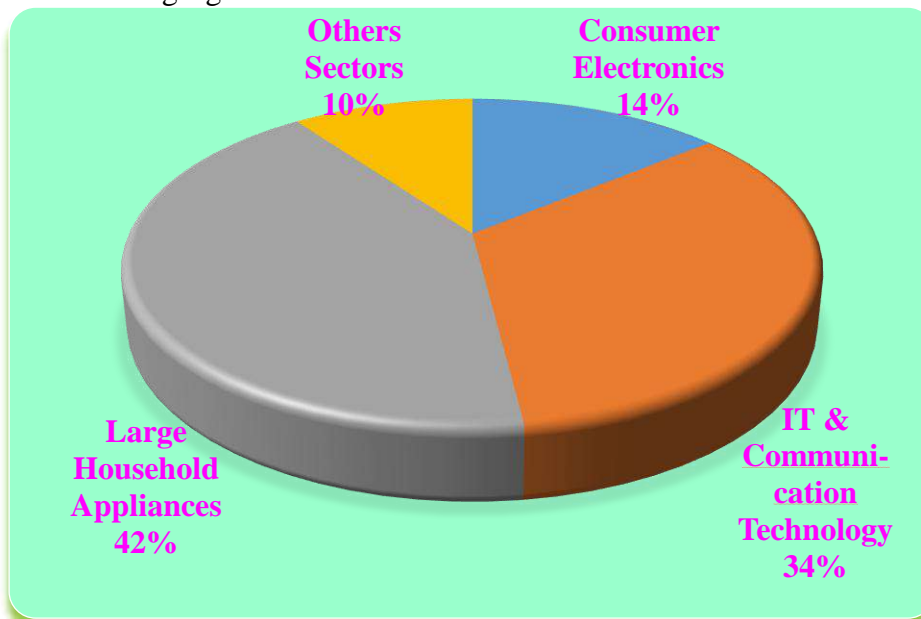


Figure-1: Sector wise percentage of collected e-waste

E-waste management strategy may include following suggested methods:

- 1) Every manufacturer, dealer & user should adapt proper way of disposal.
- 2) Collect e-waste & send it to Distributor, Dealer or Manufacturer.
- 3) E-waste can be send to CSC (Common Services Canters).
- 4) In India, total 350<sup>+</sup> CSC are available for such a dedicated purpose.
- 5) 'Digital India' campaign also adapted 'Digital Cleaning' concept.
- 6) Recycling tasks should be carried out in eco-friendly manners.
- 7) Producer should also get-back or by-back used or e-waste products.

### 4) Legislation and Policies:

As per new e-waste policy producers, dealers and distributors are abide by their mandatory e-waste EPR (Extended Producer Responsibility). There are stipulated international agreements and frameworks for scientific disposal of such materials. Also, the national regulations and enforcement guidelines has to follow at every stage of collection and recycling process.

For reuse and refurbishment of e-waste, the complete chain is promoted by the circular economical distributions. Utmost care and steps are taken for extending the lifespan of electronic appliances and devices. Public awareness, education and outreach programs are to be conducted for promoting responsible consumer behaviour.

### 5) Case Studies and Future Perspectives:

Successful E-Waste Management Programs has to be organized and conducted within every country with effective man power and management. Corporate initiatives has to be promoted for responsible recycling and disposal. At every stage, challenges and lessons learned strategies should be identified to avoid barriers to implementation the schemes. Learning from failed initiatives should studied and improvement should be adopted with recurrence.

---

For future perspectives, emerging technologies and innovations in e-waste recycling process should be upgraded with time span. Sustainable, renewable and eco-friendly product materials and designing policy should be implemented. Global collaborations, international cooperations and sharing best practices are needed to avoid the adverse effects of e-waste throughout the world.

**6) Conclusion:**

This research paper concludes by emphasizing the critical need for immediate action in addressing e-waste hazards. It advocates for the adoption of sustainable e-waste management practices, including stringent regulations, effective recycling methods, and public awareness campaigns. By comprehensively addressing the challenges associated with e-waste, society can pave the way for a more and more sustainable, eco-friendly and environmentally conscious future.

**7) References:**

1. "E-Waste: Implications, Regulations, and Management" by Rajib Shaw and Koichi Shiwaku.
2. "Environmental Impacts of E-Waste Recycling" by Sabaa Khan and Mohammad K. Parvez.
3. "Electronic Waste Management and Suggested Remedies to Control It", Dr. Gajanan S. Wajire; Published in International Journal of Current Science, Volume 13, 2023.
4. "Electronic Waste Management in India: A Stakeholder's Perspective" by Subramanian Senthilkannan Muthu.
5. "E-Waste Management Strategies: A Global Perspective"; A. Johnson and M. Patel. Published in the Journal of Environmental Management, 2018.
6. "Health & Environmental Impacts of Informal E-Waste Recycling: A Case Study in Developing Countries", S Gupta & R Sharma; Waste Management & Research, 2017.
7. "Policy Interventions for Sustainable E-Waste Management: Lessons from European Countries", L. Anderson and K. Brown; Published in Resources, Conservation and Recycling, 2019.
8. "Consumer Awareness and Behaviour towards E-Waste Recycling: A Survey Study", P. Smith and Q. Wang; Published in the Journal of Cleaner Production, 2020.
9. "Utilization of Green Electricity for Operation of Miniature Electronic Circuits", Dr. G. S. Wajire; International Journal of Scientific Research in Science and Technology, Volume 9, October 2021.

## 3

**Iris Detection by using Artificial Intelligence****Dr.R. D. Chaudhari,**

Associate Professor, Shri R.L.T. College of Science, Akola

**ABSTRACT**

Our main aim is to develop a secure biometric recognition system to identify individual person both Irises other than physical or behavioral characteristics. One such method is iris recognition which is one of the most secure and unique features of any person. God has created every individual with an exclusive iris pattern on this earth. The iris recognition technique consists of iris localization, normalization, feature extraction and matching. Their unique feature was extracted and given to the neural network using MATLAB Simulation for detecting the Iris. The match results shows that the individual is identified accurately both the iris of a same Person.

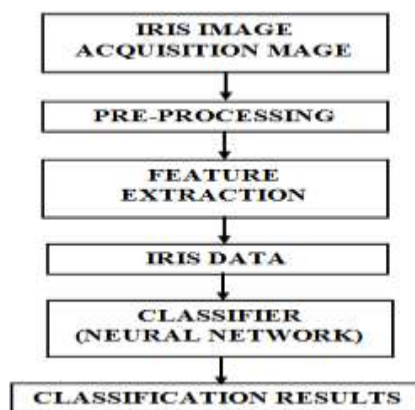
**Keywords** - Iris recognition, localization, normalization, neural network, person identification, MATLAB

**I. INTRODUCTION**

Identity verification and identification is becoming increasingly popular. Biometric measures [1] such as recognizing one's fingerprints, face, iris and voice greatly help in person identification authentication, and authorization. Pair of iris recognition has the high potential and non-invasive personal verification. Advances in the field have expanded the options to include biometrics such as iris and retina. Among the large set of options, it has been shown that the iris is the most accurate biometric. The iris is the elastic, pigmented, connective tissue that controls the pupil. Doughman [2] proposed an iris recognition system representing an iris as a mathematical function. Mayank Vatsa proposed a support-vector-machine-based learning algorithm selects locally enhanced regions from each globally enhanced image and combines these good-quality regions to create a single high-quality iris image.[3]proposes algorithms for iris segmentation, quality enhancement, match score fusion, and indexing to improve both the accuracy and the speed of iris recognition Further, Tests on another set of 801 images resulted in false accept and false reject rates of 0.0005% and 0.187% respectively, providing the reliability and accuracy of the biometric technology[5]. Leila FallahAraghi used Iris Recognition based on covariance of discrete wavelet using Competitive Neural Network (LVQ).A set of Edge of Iris profiles are used to build a covariance matrix by discrete wavelet transform using Neural Network.[4]. Today with the development of Artificial Intelligence algorithm, Iris recognition system may gain speed, hardware simplicity, accuracy and learning ability. The experimental results have shown the effectiveness of the proposed system in comparison with other previous Iris recognition system.

**A. Proposed Work Flow Chart**

The complete iris recognition system consists of 4 stages, they are image acquisition, Pre-processing, Feature extraction and Matching. Figure 3.1 shows the flow diagram of Iris recognition system.



Flow Diagram of Iris Recognition Systems

### B. Image Acquisition

This is very first step of the entire process. When a person wishes to be identified by iris recognition system, his/her eye is first photographed. The camera can be positioned between three and a half inches and one meter to capture the image. Today's commercial iris camera typically used infrared light to illuminate iris without causing harm or discomfort to the subject. In the manual procedure, the user needs to adjust the camera to get the iris in focus and needs to be within six to twelve inches of the camera. This process is much more manually intensive and requires proper user training to be successful. We must consider that the occlusion, lighting, number of pixels on the iris are factors that affect the imagequality. Figure 1.2 shows the sample iris images.

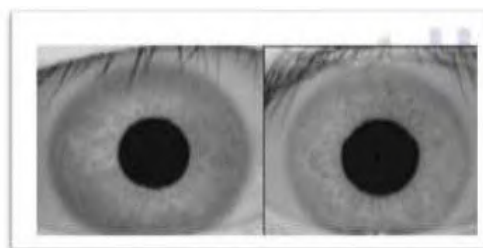


Fig. .1.2 Sample Iris Image

### C. Pre-Processing:

The acquired iris image has to be pre- processed to detect the iris, which is an annular portion between the pupil (inner boundary) and the sclera (out boundary). The first step in iris localization is to detect pupil which is the black circular part surrounded by iris tissues. The centre of pupil can be used to detect the outer radius of iris patterns. The important steps involved are :

1. Pupil detection
2. Outer iris localization

The Hough Transform is used for a quick guess of the pupil center and then the Integro- Differential Operator is used to accurately locate pupil and limbus using a smaller search space.

Canny Edge Detection can be used for detecting edges in the entire eye image and Circular Hough Transform for detecting outer boundary of iris by using pupil center and inner boundary of iris. Figure 3.1.2 shows the localized iris image.



Fig. .1.3 Localized Iris Image.

#### D. Normalization:

For the purpose of accurate texture analysis, it is necessary to compensate this deformation, since both the inner and outer boundaries of the iris have been detected so it is easy to map the iris ring to a rectangular block of texture of a fixed size. The Cartesian to polar reference transform suggested by Doughman authorizes equivalent rectangular representation of the zone of interest as shown in figure 1.3.

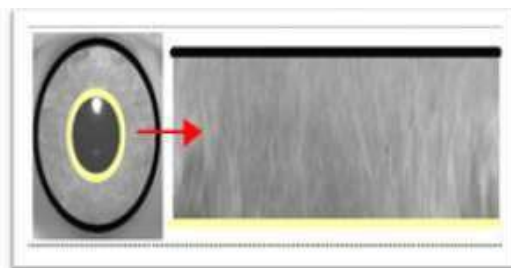


Fig. .1.3 : Iris Rectangular Representation

Pre-processing is a step, which is performed to obtain iris from the eye image. But here we have used standard iris database (i.e. UBIRIS database) so we have extracted features directly using the images.

In order to provide an accurate recognition of an individuals, the most discriminating information present in an iris pattern has been extracted. Only the significant features of the iris have been encoded so that comparison between templates is done. Below figure 1.4 shows the feature extraction stages.

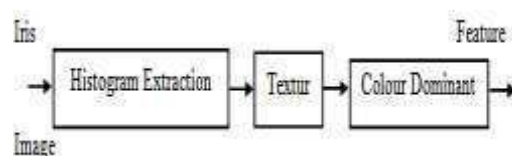


Fig. 1.4: Feature Extraction Stage

#### E. Feed forward Network:

Feed forward networks consist of a series of layers. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network's output.

Feedforward networks can be used for any kind of input to output mapping. A feedforward network with one hidden layer and enough neurons in the hidden layers, can fit any finite input – output mapping problem.

trainlm is a network training function that updates weight and bias values according to Levenberg – Marquardt optimization as shown in figure 1.5

trainlm is often the fastest backpropagation algorithm in the toolbox, and is highly recommended as a first – choice supervised algorithm, although it does require extra memory

than other algorithms.

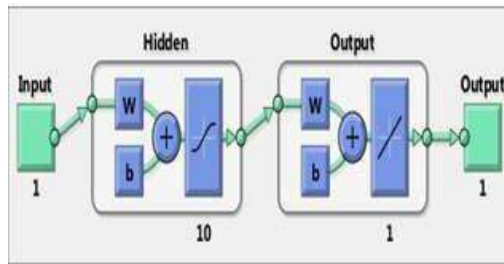


Fig. 1.5 Feedforward Network

**F. Training of Neural Network for Right Iris:**

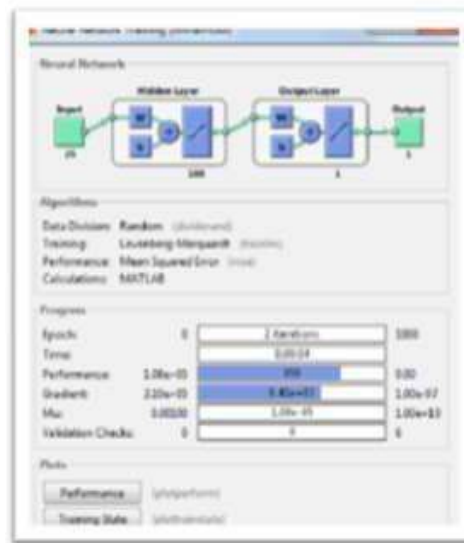


Fig. 1.6 Results of training Network for Right Iris

If the train button is pressed on the menu the neural network training (nntraintool) would be activated from the neural network toolbox. The result of train network is show in fig. 1.6.

In this figure the neural network algorithm would be displayed with 25 input two layers with weight and bias. Hidden layer is 100 and one output layer. According to the present result of training system the epoch is 2 iterations for 100 epochs. Running time is 0.004 hours. The performance is 858 for  $1.08e + 05$  target. The gradient is  $8.41e + 03$  for  $1.00e - 07$  and validation check is 0 for 6 must be displayed on to command window. According to the fig. 1.6 the neural network training system has been accomplished and known by the user neural network toolbox is very useful to simulation of this right iris recognition.

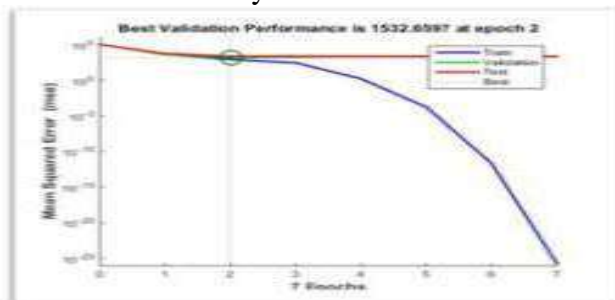


Fig. 1.7 Mean square error of Feedforward NeuralNetwork

The results are found by the algorithm and we can get the number of epochs used and which epoch gives the best result as shown in figure 1.7. As shown in fig. a plot of epoch



MSE has been plotted. The epochs get the best validation performance at epoch no. 2. The MSEW is the lowest at this point and hereafter no significant changes take place and no further decrease takes place. Hence this is the best validation performance is 1532.6597 at epoch 2. As shown in fig. 1.8. The training data are shown in the blue colour, validation is shown in blue colour, test data is shown in red colour and zero error is shown from the histogram.

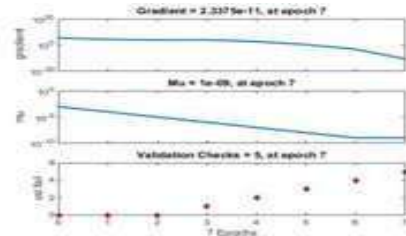


Fig. 1.8 Feedforward Neural Network Gradient versus Epoch

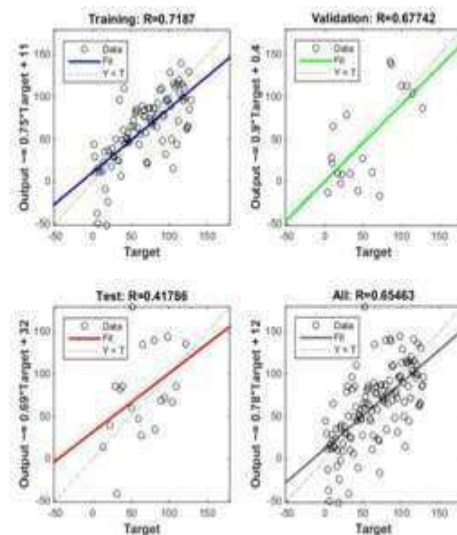


Fig.1.9 Regression plots for training, testing and validation of ANN

The four plots represent the training, validation, and testing data. The dashed line in each plot represents the perfect result – outputs = targets. The solid line represents the best fit linear regression line between outputs and targets. The R value is an indication of the relationship between the outputs and targets. If  $R = 1$ , this indicates that there is an exact linear relationship between outputs and targets. If R is close to zero, then there is no linear relationship between outputs and targets. For this example, the training data indicates a good fit. The validation and test results also show R values that greater than 0.9. The scatter plot is helpful in showing that certain data points have poor fits. As shown in figure.1.9.

## II. CONCLUSION

The proposed methodology uses Levenberg – Marquardt feedforward network. Trainlm is often the fastest backpropagation algorithm in the toolbox. According to the fig. 1.6 the neural network training system has been accomplished and known by the user neural network toolbox is very useful to simulation of this right iris recognition. The results are found by the algorithm and we can get the number of epochs used and which epoch gives the best result as shown in figure 1.7. In this paper gives the training of neural network for detection of right iris from the used database.

---

### III. REFERENCES

#### IV.

- [1]. Delac, K., &Grgic.M. "A survey of biometric recognition methods." In 46th international symposium electronics in marine, ELMAR –2004, zadalcroatia. [1] J. Daugman, "New Methods in Iris Recognition" IEEE Trans. Systems, Man, and Cybernetics, Part B, vol. 37, no.5, 2007, pp. 1167-1175.
- [2]. M. Vatsa, R. Singh, A. Noore (2008), "Improving Iris Recognition Performance Using Segmentation, Quality Enhancement, Match Score Fusion, and Indexing," IEEE Transactions on Systems, Man and Cybernetics. Part B. Cybernetics Vol. 38, N°4, pp. 1021-1035, 2008.
- [3]. Leila FallahAraghi, "IRIS Recognition Using Neural Network," Proceedings of International Multi Conference of engineers and Computer Scientist Vol I, 2010
- [4]. Bhawna chouhan, "Iris Recognition System using canny edge detection for Biometric Identification" International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 1 Jan 2011.
- [5]. Prof., S. R. Ganorkar "Person Identification using Iris Recognition" International Journal of Engineering and Technology Vol.3 No.1 February 2011.

## Design of Node MCU based Data Logger System

**R.K.Parate<sup>1</sup>, K.M.Dhole<sup>2</sup> and S.J.Sharma<sup>3</sup>**

<sup>1</sup>Department of Electronics, S.K. Porwal College Kamptee-441001

<sup>2</sup>Department of Computer Science, S.K. Porwal College Kamptee-441001

<sup>3</sup>Department of Electronics and Computer Science, R.T.M. Nagpur University, Nagpur-440033

<sup>1</sup>rkparate@yahoo.co.in

### Abstract:

Data logger is an important realization of any measurement and instrumentation system. The proposed data logger can be used in wide range of applications such as wearable and health monitoring, home automation, agriculture and farming, environmental parameters monitoring, smart energy management and IOT based applications. The designed data logger system has been tested by measuring physical parameters such as temperature, humidity and pressure. ESP32 microcontroller system is used to record data from respective sensor and the measurements are uploaded into Microsoft (MS) excel worksheet directly from Node MCU using parallax data acquisition (PLX-DAQ) software. Serial Communication between Node MCU ESP 32 and Laptop/PC is achieved using UART bus. Data obtained from the system are stored directly in excel spreadsheet which can be later used for processing and analyzing. Designed system is simple, low cost, real time that allow user to collect, monitor and plot data in excel spreadsheet.

**Keywords: Data logger, Serial communication, Node MCU ESP-32, PLX-DAQ**

### 1. Introduction

Climate plays an important role in human life. Unexpected growth of industries and vehicular traffic has seriously affected the purity of clean air and environment [1]. Temperature, Humidity and Pressure are the basic parameters that represent the environmental condition in various region of the world [2]. The conventional method of recording these data is very tedious and it has possibility of missing valuable information [3]. Continuous monitoring of these parameters and store it for future use is an important step of data logger system. The processes to collect, analyze and store data for later use is called logging. Human brain and its memory is the best data logging mechanism. When there is need to collect information faster than human brain with better accuracy then such a data logger system is useful. It acts as bridge between analog and digital environment [4,5]. It plays a major role in research and engineering applications [6]. Sensors, processor, data acquisition software, PC/Laptop and communication link are the basic blocks of data acquisition system. A successful DAQ can be made by operating all these components together. PLX-DAQ software tool help to design system that provide excel spreadsheet. It allows any microcontroller to acquire sensor data and put them in excel column as they arrive [7]. This software tool is useful for laboratory analysis of acquired sensors data and real time monitoring without need for reprogramming the microcontroller unit [8].

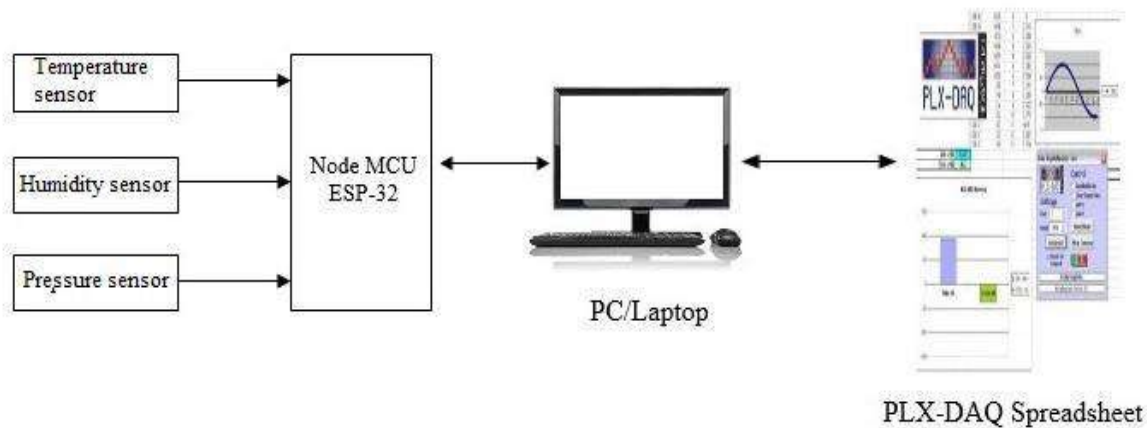
In the present work, data logger system for sensing environmental parameters is designed in our laboratory. BME 280 sensor is interfaced to Node MCU. Temperature, Humidity and Pressure are acquired from the sensor and uploaded into Microsoft excel spreadsheet using Node MCU and parallax data acquisition (PLX-DAQ) software.

### 2. Experimental

In the present work, Node MCU ESP 32 is used for processing, logging and uploading of sensor data acquired from an environment. Temperature, humidity and pressure have been

sensed using BME 280 sensor. PLX-DAQ is used for real time monitoring of data acquired using respective sensors. The proposed method of real time instrumentation is made using PLX-DAQ software tool. It consists of excel spreadsheet in which acquired data can be directly uploaded in real time into Microsoft excel (PLX-DAQ). The PLX-DAQ excel can acquire up to 26 channel of data from microcontroller.

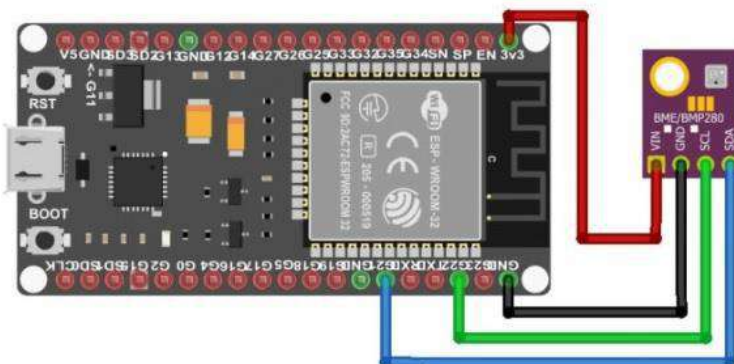
Figure 1 shows the functional block diagram of developed system in the present work.



**Figure 1: Block Diagram of the System**

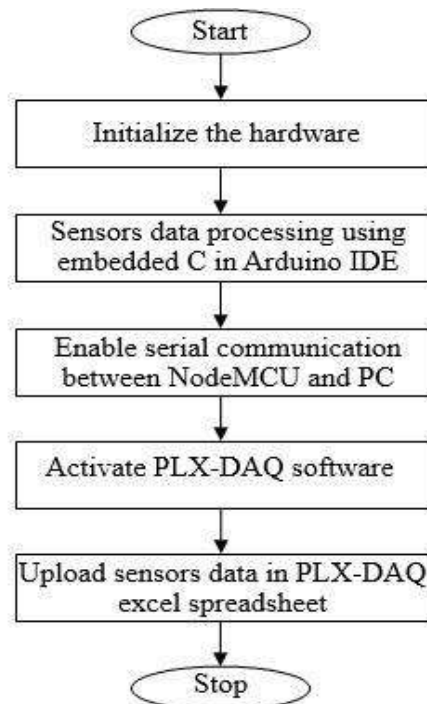
System has been implemented using hardware and software section. Hardware includes data acquisition using sensor and interfacing of sensor with Node MCU. Software section includes code development for processing of sensor information and preparation of PLX-DAQ software for uploading of sensor data directly into MS excel.

Hardware has been constructed using microcontroller unit and sensors, particularly; Node MCU (ESP32) and BME 280 sensor. Node MCU has 30 GPIO pins. Necessary signal conditioning circuits have been designed for providing outputs of respective sensor to node MCU. BME 280 provides feature of I<sup>2</sup>C interfacing having SCL and SDA. These SCL and SDA pin are connected to GPIO 22 and GPIO 21 pins of Node MCU via signal conditioning unit. Interfacing of BME 280 with ESP 32 module is shown in figure 2.



**Figure 2: Interfacing of BME 280 with ESP 32 Module**

Software includes processing of sensor data. Code has been developed in Arduino IDE using C language. Necessary libraries such as for processing sensor data have been downloaded from online sources. After installing the libraries and making circuit connection, source code is deployed into Node MCU through standard USB port by selecting proper COM port. Figure 3 shows the flowchart of developed prototype.



**Figure 3: Flowchart of System**

PLX-DAQ software tool has been used for data acquisition from Node MCU to an excel spreadsheet. This software is open source and has been downloaded from google source. After installation of this software, a folder with name PLX-DAQ automatically created on the PC/Laptop. Inside the folder PLX-DAQ, a folder with shortcut named "PLX-DAQ spreadsheet" is present. An excel sheet is open using this shortcut. Macro (a small program) have been disabled for security purpose. It might be contain virus that be harmful for other files on PC/Laptop, so Excel does not allow running macro by default. After enabling macro, control box of PLX-DAQ window come on the excel spreadsheet. This is beneficial to make connection and start excel. Various options such as baud rate, COM port selection are provided in control box for uploading data in excel sheet. In the present work, baud rate is chosen as 9600 which is same as that of Serial.begin command in Arduino IDE. Node MCU board has been interfaced with excel using PLX-DAQ software. This has been done by uploading code in Node MCU using Arduino IDE. After uploading the code, PLX-DAQ spreadsheet icon record sensor information automatically.

### 3. Results and Discussion

The test results of the data logger system have been successfully observed on PLX-DAQ software. The measured parameters are temperature, humidity, air pressure are recorded along with date and time stamp. The sensor data is also display on serial monitor in Arduino IDE software.

The designed system gives the real time data with graph of the monitored parameters with time stamp without human interference. This replaces the traditional manual recording method which posses the chance of missing valuable data. Designed system is low cost, save time and man power. The system satisfies the need of real time monitoring on a large scale that can be used to monitor environmental parameters.

---

**References**

- [1] A F Pauzi and M Z Hasan, "Development of IoT Based Weather Reporting System", *IOP Conf. Series: Materials Science and Engineering* 917 (2020) 012032
- [2] J. L. Sanchez-Cuevas, L. J. Alvarado-Galindo, S. O. Ramirez-Tobias, O. SanchezJacome, K. Cabrera-Chagoyan and L. J. Mona-Pe, "IOT Node for Monitoring Meteorological Variables Through Lorawan Technology", *World J. Engg. Res. Tech.*, 9(5), 1-8,(2023)
- [3] R. K. Parate, K.M.Dhole and S. J. Sharma, "Design of a Low Cost Biomedical Signal Acquisition System using Node-MCU ", *Int. J. Analytical and Experimental Modal Analy.*,14(8), 379-384,(2022)
- [4] P. Sarma , H. K. Singh and T. Bezboruah, "A Real-Time Data Acquisition System for Monitoring Sensor Data", *Int. J. Comp. Sci. & Engg.*, 6(6), 539-542, (2018)
- [5] Ichwana, I. S. Nasution , S. Sundari , and N. Rifky, "Data Acquisition of Multiple Sensors in Greenhouse using Arduino Platform", *IOP Conf. Series: Earth and Environmental Science*, 515 (2020)
- [6] D. T. Makusha, "Low Cost Data Acquisition System using Arduino and C", *Int. J. Res. Comp. Appl. & Robotics*, 6(1), 27-31, (2018)
- [7] O. Soe, M. Win and M. Aye, "Direct Data Uploading to Microsoft Excel by Using Parallax Data Acquisition Software and Arduino", *Univ. J. ICT Multidis. Issu. Arts & Sci., Engg, Eco, & Edu.*, 1, 305-308
- [8] A. Hammoumi, S. Motahhir, A. Chalh and A. Ghzizal and A. Derouich, "Low-cost Virtual Instrumentation of PV Panel Characteristics using Excel and Arduino in Comparison with Traditional Instrumentation", *Renewables: Wind, Water, and Solar*, 5(3),2018
- [9] [https://www.espressif.com/sites/default/files/documentation/esp32\\_datasheet\\_en.pdf](https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf)
- [10] [www.arduino.cc](http://www.arduino.cc)
- [11] [www.randomnerdtutotrils.com](http://www.randomnerdtutotrils.com)

## 5

## Exploring Innovations in Sensor Technologies for Enhanced Healthcare Monitoring Systems: A Comprehensive Analysis

**K. Y. Rokde<sup>1\*</sup>, S. S.Shende<sup>2</sup>**

<sup>1</sup>Assistant Professor, Department of Electronics, Dr. Ambedkar College, Nagpur, India <sup>1</sup>

<sup>2</sup>Assistant Professor, Department of Electronics, Sharadchandra Pawar College, Lonand, India <sup>2</sup>

Email: [krokde4@gmail.com](mailto:krokde4@gmail.com)

### ABSTRACT:

This paper presents a comprehensive analysis of exploring innovations in sensor technologies for enhanced healthcare monitoring systems. The prevalent health issues faced by individuals today pose significant challenges, and diverse solutions are available. However, employing sensors for real-time health monitoring offers a proactive approach, allowing for the continuous assessment of crucial parameters. By analyzing these parameters, timely solutions can be identified. Healthcare Wireless Sensor Networks (HWSNs) play a pivotal role in facilitating access to these sensors, enabling continuous monitoring of patients. This specialized field ensures ongoing connectivity to patients' sensors, allowing the system to promptly detect and alert medical professionals to any changes in the patient's condition. The integration of HWSNs contributes to enhanced patient health monitoring in a seamless manner.

In this paper, healthcare monitoring is streamlined through the utilization of MATLAB software for disease diagnosis. This approach enables the system to provide instant precautions to patients effortlessly and administer targeted treatments, particularly in critical conditions.

**Keywords** - *Sensors, Healthcare, HWSNs, MATLAB.*

### I. Introduction

#### 1.1 Healthcare Wireless Sensor Network [HWSNS]

The application of sensor networks in healthcare is a specialized domain at the intersection of electronics, computing, and the internet. This system, operating with electronic devices, collects and wirelessly transmits data globally. While applicable to various fields such as forests, traffic, and environmental monitoring, its significance is most pronounced in medical science, where it plays a vital role in human life. In hospitals, traditional monitoring involves periodic checks by staff, limiting the acquisition of parameters to specific intervals. However, the implementation of Healthcare Wireless Sensor Networks (HWSNs) enables continuous data retrieval without the need for direct patient visits. This is particularly beneficial in critical situations, allowing staff to provide timely attention using sensors attached to patients. The integration of sensors into HWSNs enables remote data monitoring through the internet [1].

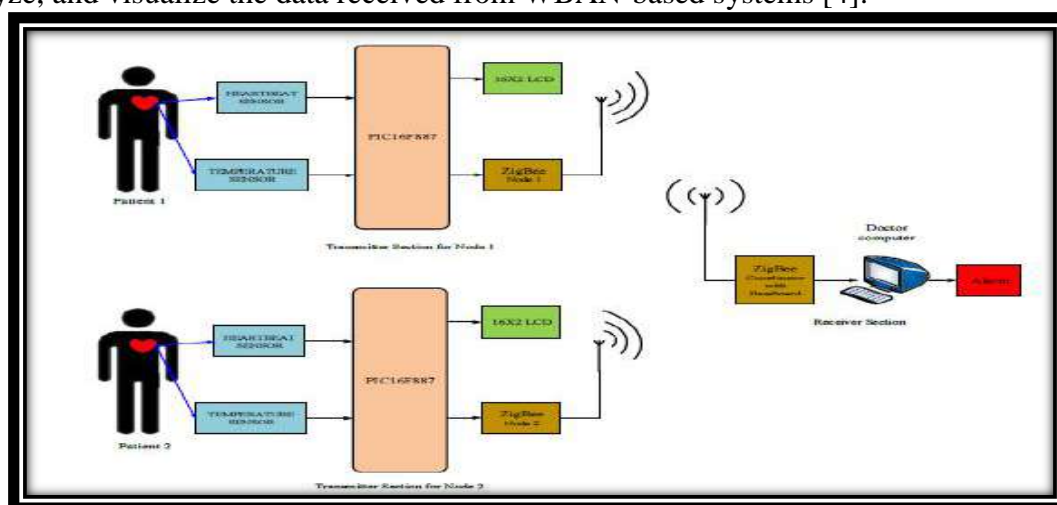
One notable challenge in employing sensor nodes is ensuring patient mobility, especially during hospitalization, to enhance their overall well-being. This poses a task for HWSNs to support the mobility of sensor nodes as permitted by patients. Addressing network coverage issues related to sensor node mobility, HWSNs need to expand multiple access ports and accommodate route variations to reach each sensor node [2]. Continuous access to sensor nodes requires a valid route available at all times. Handover mechanisms, facilitating the transition of power attachment changes within the network, become crucial. The effectiveness of the handover mechanism significantly influences continuous connections to sensor nodes in HWSNs. This paper emphasizes recent handover mechanisms that specifically support sensor node intra-mobility, requiring close control and timely data acquisition [3].

The fundamental principles of HWSNs encompass real-time monitoring, the random and continuous motion of sensor nodes, and the desire for prolonged battery life.

## 1.2 Wireless Body Area network

Wireless Body Area Network (WBAN) stands as a specialized category within biomedical sensor networks, where the biomedical sensor nodes are strategically placed on, near, or within the human body. Within the realm of medical healthcare systems, WBAN plays a pivotal role in continuously monitoring the health of individuals, especially the elderly or those facing health challenges. The biomedical nodes within WBAN are responsible for sensing and processing vital signs, including heart rate, blood pressure, body temperature, and respiratory metrics, derived directly from the human body. Subsequently, this collected data is transmitted to a medical center through a base station, allowing medical professionals to monitor human health remotely.

In the medical center, doctors and caregivers rely on monitoring systems/interfaces to process, analyze, and visualize the data received from WBAN-based systems [4].



**Fig 1. Block diagram of Functional System**

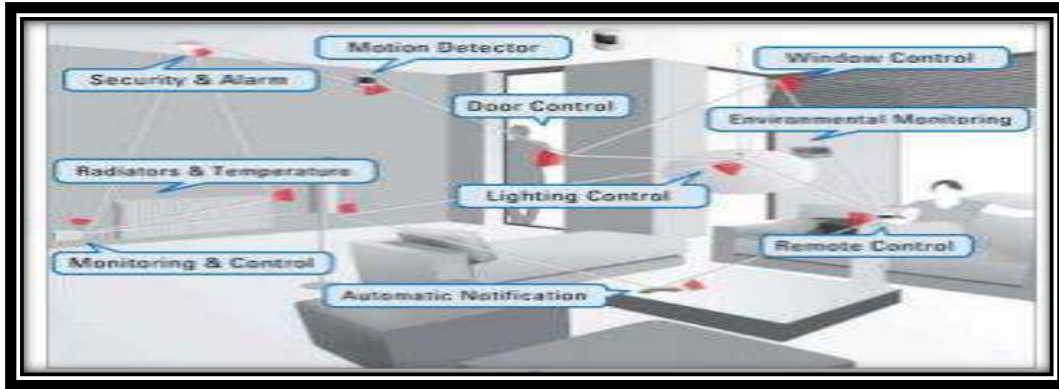
These systems leverage wireless technologies to transmit vital signs for medical evaluation, primarily serving as a means to transfer monitoring information about patients in hospitals. Wireless sensor networks employ various technologies, including Infrared, Bluetooth, ZigBee, Touchscreen, among others. Despite the angle limit challenge in Infrared transmission and its limited use for physiological signal transmission, Bluetooth, and ZigBee stand out as prominent technologies. While Bluetooth boasts a superior transmission rate, ZigBee is preferred due to its significantly lower power consumption. Consequently, ZigBee is widely utilized for continuous 24-hour monitoring in communication transmission systems. In comparison to Bluetooth, ZigBee offers enhanced network flexibility, supports a larger number of nodes, provides better transmission range, and operates with low power consumption. The ability to accommodate a large number of nodes facilitates the scalability and expansion of these systems. Recent applications have tested ZigBee-based wireless networks in various scenarios, showcasing their effectiveness in different healthcare contexts [5].

Zigbee, a short-distance wireless transmission technology, has experienced rapid development in recent years. Established by the Zigbee Alliance and the IEEE802.15.4 Team in collaboration with various companies, Zigbee prioritizes low costs and low power consumption. Originally designed for general home networks, its swift evolution has extended its application into diverse industries, including medicine. Zigbee boasts a low transmission rate (250 Kbps), short-range capability (typically 50m-100m), and low power consumption, making it easily deployable and adaptable. The technology supports a substantial number of network nodes and accommodates various network topologies. Additionally, Zigbee is recognized for its speed, reliability, and cost-effectiveness. Its affordability makes it a popular



choice for simple wireless control applications in home industries and medicine, such as toxic gas (carbon-monoxide) detection, medical sensors, and patient emergency alarm systems.

The inherent advantages of Zigbee, including its ability to transmit signals at any time, as well as its wireless positioning function, contribute significantly to medical treatment procedures. In comparison to traditional wired equipment, Zigbee enhances healthcare applications by enabling real-time monitoring of patient movements, body temperature, or pulse at any given moment [6].



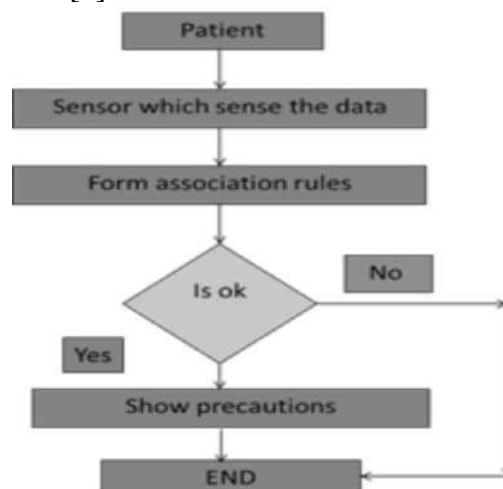
**Fig 2. Zigbee Technology and its application**

## II. Objective of Research Work

The primary aim of this research is to simulate an intra-mobility solution, with a specific focus on Healthcare Wireless Sensor Networks (HWSNs). In hospitals, patients often experience prolonged waiting times to consult with doctors, and in critical conditions, waiting becomes impractical, leading to various challenges. The central goal of this study is to leverage software to empower patients in diagnosing diseases and receiving instant precautions effortlessly, eliminating the need for additional efforts or means. Through this approach, patients can efficiently communicate their health concerns, receive timely disease-related precautions, and seek solutions [7].

## III. System Configuration

The simulation model utilizes MATLAB for its high-performance computation and visualization capabilities. A key strength lies in its user-friendly interface, making it easily accessible and manageable. Users can input their symptoms, such as temperature, blood pressure, ECG, and heartbeats, through sensors. In return, the system provides predictions based on the user's symptoms [8].



Flowchart for procedure:

**Fig 3. Flowchart of System**

## System overview

These systems consist of four essential components:

1. Sensing Module: This module enables the collection of specific parameters.
2. Processing Module: Comprising a microcontroller, this module assesses the capacity of the sensor node to execute programs and process data [9].
3. Communication Module: Responsible for wirelessly transmitting data to a network, typically adhering to the IEEE 802.15.4 standard.
4. Power Supply: Serving as a crucial element, the power supply functions as the energy source to sustain the node's operation [10].

## IV. Conclusion

In this paper, we present a wireless health monitoring system which is able to receive data on receiver side and immediate action will be taken according to the results obtained. The system offers an effective solution for enhancing the current health infrastructure by integrating various sensors onto a single platform. Leveraging Healthcare Wireless Sensor Networks (HWSNs), this system enables a comprehensive analysis of a patient's disease, significantly reducing the time required for diagnosis. The system delivers textual information to the patient, allowing them to access and act upon it from anywhere. Its convenience and efficiency foster increased interaction between patients and doctors, ultimately minimizing unforeseen tragedies. Employing this technique for patient monitoring proves to be a crucial aspect of healthcare precaution. The integration of this technology has the potential to enhance healthcare services and outcomes [11].

## V. Application of WSN in Healthcare

Healthcare is a perpetual concern, given its direct impact on an individual's quality of life. The emphasis on preventive measures over reactive treatments underscores the importance of regular individual monitoring. The aging demographic in developed countries poses a substantial challenge to government budgets, particularly with an increasing number of elderly individuals residing in independent senior housing. Traditional healthcare practices involve periodic check-ups, relying on patients to recall symptoms, subsequent diagnostic checks by doctors, and ongoing monitoring of treatment progress if applicable [12]. However, certain symptoms may only manifest during daily activities, making continuous monitoring crucial. Wireless sensor networks in healthcare applications offer solutions such as in-home assistance, smart nursing homes, and augmentation for clinical trials and research. In-home healthcare becomes imperative for conditions like Parkinson's or Alzheimer's, providing memory enhancement through medicine reminders, mental stimulation via sounds or images, location control of home appliances, access to medical data, and swift response to emergency situations. This approach suggests a potential multi-tiered architecture, involving light weight mobile computers and smart sensors in conjunction with more powerful computational devices [13].

## Healthcare Application

1. Glucose level monitoring
2. Asthma.
3. Preventing medical accidents.
4. Cardiovascular diseases
5. Alzheimer, depression, and elderly people monitoring.
6. Stroke and post-stroke.
7. Artificial retina.
8. Home monitoring
9. Heart rate monitoring

10. Artificial retina.
11. Mobihealth
12. Multi-electrophysiological system[14].

## References

- [1]Jung, J.Y., Lee, J.W.: Improved WBAN Communication at Mental Healthcare System with the Personalized Bio Signal Devices. In: The Proceedings of 8th International Conference Advanced Communication Technology, Korea, pp. 812–816 (2006).
8. Vaudenay, S.: On the need for Multipermutations: Cryptanalysis of MD4 and SAFER. In: Preneel, B. (ed.) FSE 1994. LNCS, vol. 1008, pp. 286–297. Springer, Heidelberg (1995).
- [2]U. Anliker, J. A. Ward, P. Lukowicz, G. Tröster, F. Dolveck, M. Baer, F.Keita, E. Schenker, F. Catarasi, and R. Schmid, “AMON: A Wearable multiparameter medical monitoring and alerts system”, IEEE Trans. On Inf.Technol., Biomed., vol. 8, no. 4, pp. 415-427, 2004.
- [3] Istepanian, R.S.H., Jovanov, E., Zhang, Y.T.: M-Health: Beyond Seamless Mobility and Global Wireless Health-Care Connectivity. The Proceedings of the IEEE Transactions on Information Technology in Biomedicine, 405–414 (2004).
- [4] Hairong Yan, Hongwei Huo, “Wireless Sensor Network Based E-Health System-Implementation and Experimental Results”, Member, IEEE Trans. On Consumer Electronics, vol. 56, no. 4, November 2010.
- [5] Aung Soe Phyo, Zaw Myo Tunand, Hla Myo Tun, “Wireless Patient Monitoring System using point to multipoint Zigbee Technology”, ISSN:2277-8616, Vol. 4, Issue(6), June 2015, pp 267-274.
- [6] Anupam Jain, Minakshi Halder, “An overview of wireless body area network[WBAN] using Zigbee Technology”, ISSN: 2455-2631, Vol. 1, Issue [5], May 2016, pp 888-895.
- [7]S. M. Ghatole, K.Y. Rokde, S. S. Shende, P.B. Dahikar, “Role of Wireless Body Area Network in Remote Healthcare Monitoring” published in International Journal of Researches in Biosciences, Agriculture and Technology (IJRBAT), ISSN: 2347-517X, Volume II, issue (7), Nov 2015, pp 154-157.
- [8] Chin-Teng Lin, Fellow, Kuan-Cheng Chang, Chun-Ling Lin, Chia-ChengChiang, Shao-Wei Lu, Shih-Sheng Chang, Bor-Shyh Lin, Hsin-Yueh Liang,Ray-Jade Chen, Yuan-Teh Lee, and Li-Wei Ko, “An IntelligentTelecardiology System Using a Wearable and Wireless ECG to Detect AtrialFibrillation” Member, IEEE Transaction on Information technology inbiomedicine, Vol. 14, No. 3, May 2010
- [9]K. Y. Rokde, P. B. Dahikar, M. J. Hedau, S. M. Ghatole, S. S. Shende “Study of Biosensors using nanotechnology” published in International Journal of Advances in Science, Engineering and Technology (IJASEAT), ISSN: 2321-9009, Special Issue-1, June- 2015, pp 155-157.

- 
- [10] K. Y. Rokde, S. M. Ghatole, A. G Kshirsagar, N. D. Meshram, S. S. Shende "Design and Implementation of Speed Control Motor Using Fuzzy Logic Technique" International Journal of Industrial Electronics and Electrical Engineering (IJIEEE), Volume 4, Special Issue 2, June 2015, ISSN: 2347-6982, pp 120-124.
- [11] S. M. Ghatole, K. Y. Rokde, S. S. Shende, P.B. Dahikar "Healthcare System with Interactive Biosensors" published in International Journal of Electronics, Communication & Soft Computing Science and Engineering (IJECSCE), ISSN: 2277-9477, Volume 4, Issue 4, July 2015, pp 1-4.
- [12] Chandani Suryawanshi, Bhakti Kurhade, "A WSN based System for Enhancing Intra Mobility Solution for Healthcare", International Journal of Emerging Trends in Engineering and Development (IJETED), Issue 5, Vol. 4 (June.-July. 2015), ISSN 2249-6149.
- [13] M. Alwan, S. Kell, B. Turner, S. Dalal, D. Mack, and R. Felder, "Psychosocial Impact of Passive Health Status Monitoring on Informal Caregivers and Older Adults Living in Independent Senior Housing," in 2nd Information and Communication Technologies, Surabaya, Indonesia, 2006, pp. 808-813.
- [14] J. A. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, S. Lin, S. Son, R. Stoleru, and A. Wood, "Wireless Sensor Networks for In-Home Healthcare: Potential and Challenges," in High Confidence Medical Device Software and Systems Workshop, Pennsylvania, USA, 2005.
- [15] Paulo Neves, Michal Stachyra, Joel Rodrigues, "Application of Wireless Sensor Networks to Healthcare Promotion", Journal Of Communications Software And Systems, Vol. 4, No. 3, September 2008.

---

## 6

### **An IoT Application- Wireless Water Tank Meter Using Ultrasonic Sensor**

**Miss. Teesha Das<sup>1</sup>, Miss. Aastha Orkey<sup>1</sup>**

1. Student Department of Electronics Hislop College, Civil Lines, Nagpur, Maharashtra, India

#### **Abstract**

The efficient management of water resources is a critical aspect of sustainable development. In this context, the integration of Internet of Things (IoT) technologies offers innovative solutions for real-time monitoring and management of water-related infrastructure. This project presents the design, development, and implementation of a wireless water tank meter utilizing ultrasonic sensor technology as a key component of an IoT application. The proposed system aims to provide accurate and real time information about water levels in tanks, enabling users to make informed decisions for water conservation and resource optimization.

**Keywords: Ultrasonic sensors, ESP 32 board, Blynk App, Wireless module, IoT, Real time monitoring, microcontroller, IoT.**

#### **Introduction**

Water level indicator is a must have gadget, if you own a house or even living in a rented house. A water level indicator shows one important data for your house which is as important as your energy meter's reading, that is, how much water is left? So that we can keep track of water consumption and we don't need to climb upstairs to access the water tank to check how much water left and no more sudden halt of water from faucet.

The increasing demand for water resources and the growing concern over water scarcity underscore the need for effective water management systems. IoT technologies have emerged as powerful tools for monitoring and controlling various aspects of water infrastructure. This project focuses on the development of a wireless water tank meter using ultrasonic sensors, contributing to the advancement of smart water management solutions.

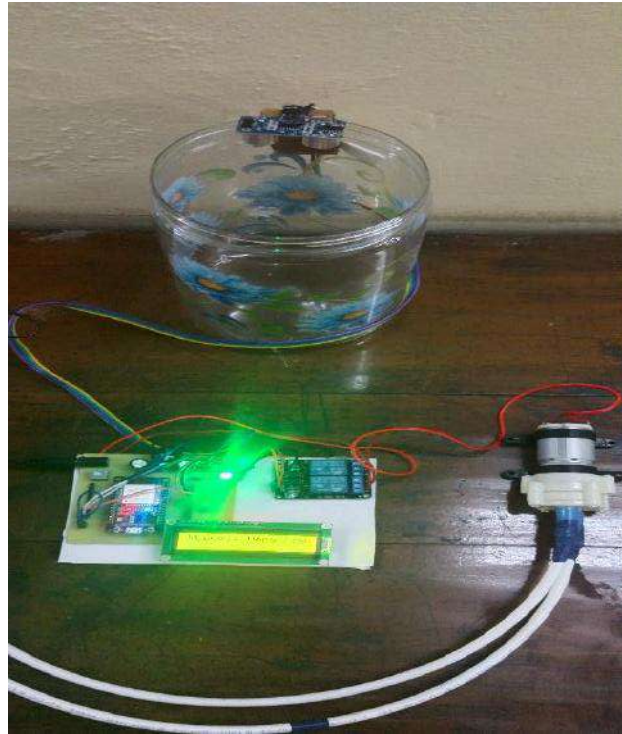
IoT-based water level monitoring provides real-time autonomous detection of water levels and takes appropriate action based on the levels including overflowing, water depletion, and water usage.

#### **System Architecture**

The proposed system comprises three main components: ultrasonic sensors for water level measurement, a microcontroller for data processing, and wireless communication for real-time data transfer to a cloud-based platform. The ultrasonic sensors utilize sound waves to measure the distance between the sensor and the water surface, providing accurate water level readings. This meter sense the water level using an ultrasonic sensor placed under the lid of the tank and sends the data via a wireless transmitter to the receiver unit.

- The LCD in the receiver unit displays the water level as a long bar.
- The water percentage is calculated and displays in the LCD
- A short message like "Low", "Medium", "Full", is shown in the display
- The receiver contains wireless device which indicates corresponds to the water level
- The receiver also has a buzzer which buzzes when the water level is too low or when the tank is full, when refilling.

- We have used Blynk mobile Application to monitor and operated wireless water tank meter.

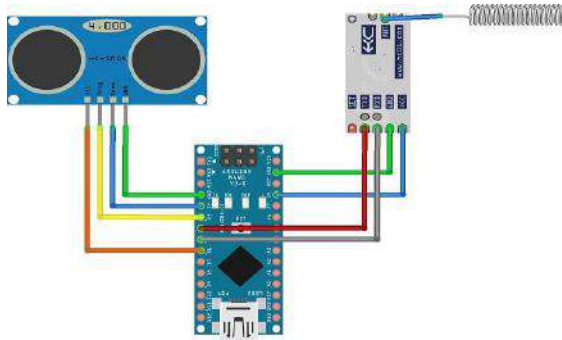


### Ultrasonic Sensor Technology

This section explores the principles of ultrasonic sensor technology and its suitability for water level measurement. The advantages and limitations of ultrasonic sensors in the context of water tank monitoring are discussed, along with potential calibration methods to enhance accuracy.

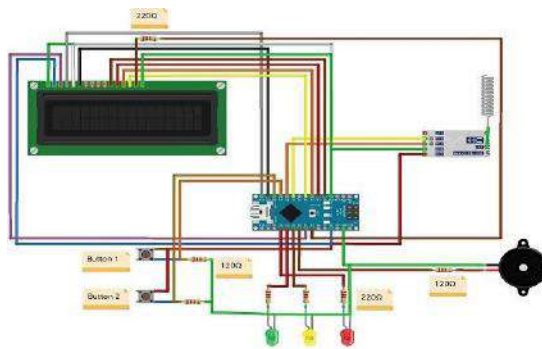


### Microcontroller and Data Processing



The microcontroller acts as the brain of the system, processing data from the ultrasonic sensors and transmitting it to the cloud platform. The paper delves into the selection criteria for the microcontroller, its role in data processing, and the development of algorithms for precise water level calculations.

### Wireless Communication



Efficient wireless communication is crucial for real-time monitoring. This section discusses the selection of wireless communication protocols and their integration into the system. The paper also explores security measures to protect the transmitted data from unauthorized access.

### Cloud-Based Platform

The collected data is stored and analyzed on a cloud-based platform, providing users with a user-friendly interface for monitoring water levels, historical data analysis, and generating alerts. The advantages of cloud computing in enhancing scalability, accessibility, and data management are discussed.

### Experimental Results

The research includes experimental results demonstrating the effectiveness and accuracy of the proposed wireless water tank meter. Real-world scenarios and performance metrics are presented, validating the reliability of the system in diverse environmental conditions.

The Testing is a process to answer the hypothesis: could the ultrasonic sensor read level of the water container and how good the accuracy? From testing the ultrasonic sensor to read water level on static condition 0cm, 10 cm, 25 cm, and 45 cm(maximal) with several timers on ping process: without a timer, timer 1ms, timer 10ms, 60ms. Data is taken from 50 data in the database Data result is:

Table 1 Comparison reading level at water container empty using the ultrasonic sensor

Comparison of the reading level at water container empty				
	Without Timer	Timer 1 ms	Timer 10 ms	Timer 60 ms

Average Error Result	35.36 cm	16.68 cm	0.02 cm	0 cm
----------------------	----------	----------	---------	------

Table 2 Comparison reading level at water container 10 cm using ultrasonic sensor

Comparison of the reading level at water container 10cm				
	Without Timer	Timer 1 ms	Timer 10 ms	Timer 60 ms
Average Error Result	1.32 cm	0.58 cm	1 cm	1 cm

Table 3 Comparison reading level at water container 25 cm using ultrasonic sensor

Comparison of the reading level at water container 25 cm				
	Without Timer	Timer 1 ms	Timer 10 ms	Timer 60 ms
Average Error Result	0.14 cm	0.02 cm	0 cm	0 cm

Table 4 Comparison reading level at water container 45 cm using ultrasonic sensor

Comparison of the reading level at water container empty				
	Without Timer	Timer 1 ms	Timer 10 ms	Timer 60 ms
Average Error Result	0.02 cm	0.04 cm	0 cm	0 cm

That's the result of the accuracy of the ultrasonic sensor on the static water level.

### Conclusion and Future Work

The research paper concludes with a summary of key findings, highlighting the significance of the developed IoT application for wireless water tank monitoring. Future work includes potential enhancements, scalability considerations, and integration with other IoT-enabled water management systems.

- From experiments, the water container at working mode can be concluded that when the pump from running to stop, the error of reading level from the sensor because it is caused by the flow of water that causes the surface of the water is not flat so the reflection pulse not perfect any more.
- From the connectivity experiments with Wi-Fi, it can be concluded that Microcontroller ESP8266 can be connected with Wi-Fi with security protocol WPA-PSK, WPA2-PSK, as well as WEP protocol.
- From experiments, ultrasonic sensor can be used as water monitoring with error threshold between 2 cm



---

**REFERENCES**

Ismail, Ammar Asyraf, Muhammad Arief Azizi, and Asnazulfadhli Zariman. "Smart Water Level Indicator." *International Journal of Recent Technology and Applied Science (IJORTAS)* 2.1 (2020): 48-58.

Amirruddin, Melaty, et al. "Microcontroller based water level indicator using GSM modem: design and application." *1st International Conference on Future Trends in Computing and Communication Technologies*. 2012.

Kumar, Ajith, Chandrakant Naikodi, and L. Suresh. "Water Level Indicator using Smart Bluetooth." *Int. Journal of Engineering Research* 5 (2016): 790-991.

Chinaeke-Ogbuka, Ifeanyi, Augustine Ajibo, and Cosmas Ogbuka. "A Microcontroller-Based Water Level Indicator using Radio Frequency (RF) Technology and Ultrasonic Sensor." *Dimension* 45.20 (2019): 15mm.

Rachel, P. Nancy, et al. "Automatic Water Level Indicator and Controller by using ARDUINO." *International Journal of Research in Engineering and Technology eISSN* (2019): 2319-1163.

Bordoloi, Dibyahash, and Surendra Shukla. "The Internet of Things (IoT)-automated smart water level indicator: a practical application of smart irrigation." *Webology* 18.5 (2021): 3201-3205.

Jeswin, C. Jestop, B. Marimuthu, and K. Chithra. "Ultrasonic water level indicator and controller using AVR microcontroller." *2017 international conference on information communication and embedded systems (ICICES)*. IEEE, 2017.

Hassan, Wan Haszerila Wan, et al. "Flood disaster indicator of water level monitoring system." *International Journal of Electrical and Computer Engineering* 9.3 (2019): 1694.

## Artificial Intelligence in Agriculture: An Analytical Study

R. J. Gajbe<sup>1</sup>, C. R. Chaudhari<sup>2</sup>, A. R. Yenkar<sup>3</sup>, N. R. Bundile<sup>4</sup>, K. D. Bhanang<sup>5</sup>

<sup>1,3,4,5</sup>- Vidya Bharati Mahavidyalaya, Amravati (MS).

<sup>2</sup>- Mahatma Fule Mahavidyalaya, Warud (MS)

<sup>1</sup>-rjgajbe20@gmail.com

<sup>2</sup>- itsmechandu04@gmail.com

<sup>3</sup>- anupyenkar1996@gmail.com

<sup>4</sup>- bundilenilesh66@gmail.com

<sup>5</sup>- kirtibhanang14@gmail.com

### Abstract:

Around 2050, the global population will reach around 10 billion. It will be placing significant pressure on the agricultural sector to increase the crop production. The researchers in the world are thinking whether the AI plays crucial role in agriculture to increase yields and develops ecosystem inspite of global challenges such as climate change, population growth together with resource scarcity threaten the sustainability of our food system. Up to some extent, the researcher found that the Artificial Intelligence works better move in agriculture sectors.

**Keywords:** Artificial Intelligence, Machine Learning, Deep Learning, Agriculture

### 1. Introduction:

According to UN Department of Economic and Social Affairs, the current world population of 7.6 billion is expected to reach 8.6 billion in 2030, 9.8 billion in 2050. It will be placing significant pressure on the agricultural sector to increase the crop production. To increase yields, it needs to expand the land use and adopt large-scale farming, or by implementing innovative practices and adopting advance technologies on existing farmland. The today's challenges of farmers to achieving desired farming productivity are- they have limited land holdings, shortages of labors, climatic changes, environmental issues, and declination of soil fertility. With the help of successful implementation of real-life technological solutions such as artificial intelligence, we may cope up these situations. Our focus has been developing innovative systems for quality control, compliance practices, and more. Now, we will dive deeper into how new technologies can help the farming business move forward.

### 2. AI Technologies used in Agriculture:

In recent years, the world has observed rapid advancements in agricultural technology, revolutionizing farming practices. These innovations are becoming increasingly essential as global challenges such as climate change, population growth together with resource scarcity threaten the sustainability of our food system. Introducing AI solves many challenges and helps to reduce many disadvantages of traditional farming.



Figure 1: Data Collection from Different Sources in Agriculture

### 2.1 Artificial Intelligence Technologies:

Artificial intelligence is a field in which with the help of computer science and robust datasets to able the machines for problem-solving and decision making. Machine learning and deep learning are the sub-fields of AI and deep learning is a sub-field of machine learning. These disciplines are comprised of AI algorithms to create expert systems which make predictions or classifications based on input data.

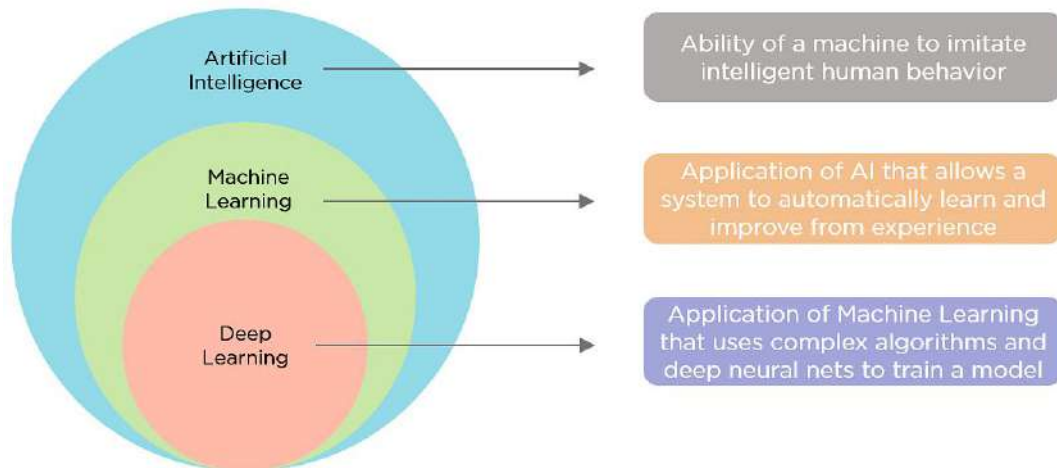


Figure 2: Differences between artificial intelligence, machine learning, and deep learning

### 2.2 Flowchart of Monitoring & Control Operations in Agriculture:

A flow chart is shown in Figure 3 shows the complete process for Agriculture field monitoring and control operations using Sensors, IoT & Machine Learning Algorithm/ Deep Learning Algorithm on Mobile Phone or PC.

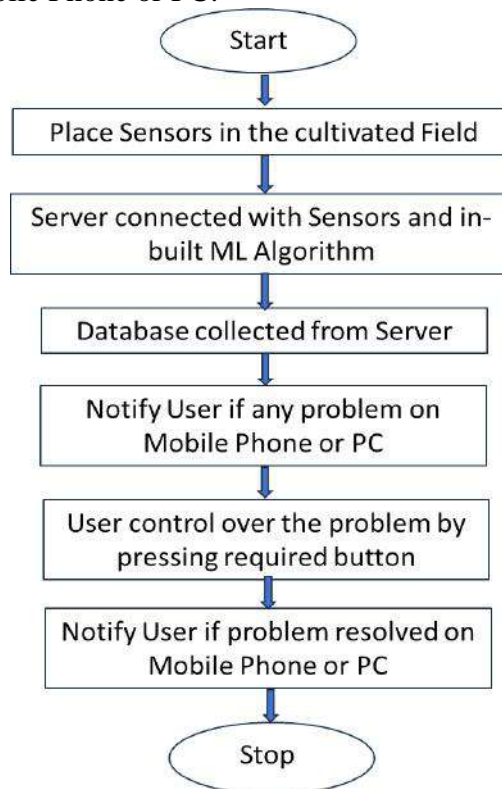


Figure 3: Flowchart of Complete Process of Monitoring and Control Operations

### 2.3 Machine/ Deep Learning Models used in Agriculture:

Machine/ Deep learning is a growing field with many potential applications in agriculture. Farmers and agricultural scientists are exploring how turning to machine learning development can improve crop yields, reduce water usage, and predict pests and diseases. In the future, machine learning may help farmers to use resources more efficiently and produce food sustainably.

**Regression models:** The Regression Model can predict crop yields, the price of agricultural commodities, and the demand for agricultural products. In addition, regression models can be used to study the impact of new technologies on agriculture and to evaluate the efficiency of different agricultural production planning systems.

**Clustering model:** The Clustering Model can be used in agriculture to group plants with similar characteristics. This can be useful for identifying which plants are more likely to thrive in certain conditions and developing targeted interventions.

**Bayesian Model:** The Bayesian Model can provide accurate crop yield predictions by using data on weather, soil conditions, and other factors. Farmers can use this data to decide what crops to plant, how much fertilizer to use, and when to harvest.

**Convolutional Neural Networks:** The CNN is a type of DL algorithm composed of multiple convolutional layers, pooling layers, and fully connected layers. CNN techniques are used in many applications of agriculture field such as to predict upcoming water needs and to provide clues that can be helpful for real-time irrigation management. It can also helpful in counting fruit using automated fruit detection and algorithms can optimize agriculture production and help in managing the harvest process effectively.

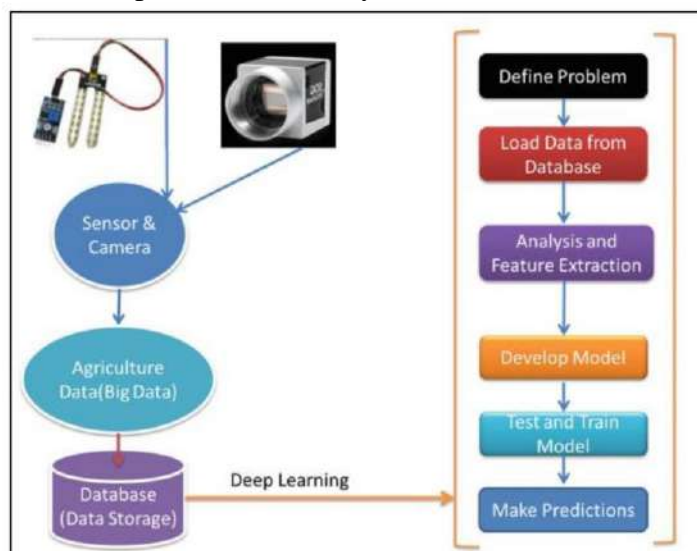


Figure 4: Deep learning approach for water management

**Artificial Neural Network:** The Artificial Neural Networks are well-suited for agricultural applications because they can learn to identify patterns in data that are too complex for humans to separate. For example, artificial neural networks can be used to develop new strains of crops that are more resistant to pests or diseases.

### 2.4 Advantages of Machine/ Deep Learning in agriculture:

1. Machine learning helps farmers optimize their irrigation schedules, fertilizer application rates, and pesticide use to reduce wastage and environmental harm.
2. ML automates field mapping, monitoring crop health, and applying fertilizers. This can save farmers time and money and reduce their need.
3. ML helps farmers to optimize resources, resulting in increased crop yields. This can help to improve food security and reduce hunger.
4. ML helps farmers save money on crucial resources like water, fertilizer, and pesticides. This can increase profitability and make farming more sustainable in the long term.
5. Machine learning enables farmers to make better decisions about when to plant, how to irrigate, and when to apply fertilizers.
6. ML helps farmers avoid hazardous tasks such as working with pesticides. This can improve farm workers' safety and health.
7. ML can provide farmers with personalized recommendation for planting, irrigation, and fertilization.
8. ML can help farmers adapt to changing conditions by identifying optimal growing conditions and developing early warning systems for extreme weather events.

Implementing machine/deep learning in agricultural processes works more efficiently by reducing environmental harm, improving yields and food quality, and making processes. Those who may adopt these technologies in future will be in well-positioned to get more benefits.

### 3. Conclusion

AI in agriculture offers numerous opportunities to farmers, including improved crop health monitoring, precision farming, and weather forecasting. However, farmers face several challenges when adopting AI, including the cost of implementing AI systems and the need for technical expertise. The expertise from various companies provides comprehensive services, including micro-financing and AI-based systems, to help farmers harness the power of AI and reach their full potential. Government may also provide subsidy facility and their expertise to the farmers in regards with automation towards farming. Thus, the farmers may improve crop health, optimize operations, and increases profitability, paving the way for a brighter future in agriculture.

### References:

1. Dhruv Sawhney (2024), Machine Learning In Agriculture: 13 Use Cases & Examples, *itransition Machine Learning*; <https://www.itransition.com/machine-learning/agriculture>, Pg. 1–15.
2. Gardezi, Joshi, Rizzo, Ryan, Prutzer, Brugler & Ali Dadkhah (2023), Artificial intelligence in farming: Challenges and opportunities for building trust; *Agronomy Journal*, Pg. 1–12.
3. Himanshu Sharma, Ahteshamul Haque & Zainul Abdin Jaffery (2019); Smart Agriculture Monitoring using Energy Harvesting Internet of Things (EH-IoT); *An International Scientific Journal/ World Scientific News 121* (2019) 22-26.
4. Jiali Zha (2020), Artificial Intelligence in Agriculture; *Journal of Physics: Conference Series*, (IOP Publishing), doi:10.1088/1742-6596/1693/1/012058
5. Kirtan Jha, Aalap Doshi & Poojan Patel (2018); Intelligent Irrigation System using Artificial Intelligence and Machine Learning: A Comprehensive Review; *International Journal of Advanced Research/ Int. J. Adv. Res.* 6(10), 1493-1502, DOI URL: <http://dx.doi.org/10.21474/IJAR01/7959>.

6. Maha Altalak, Mohammad Ammaduddin, Amal Alajmi & Alwaseemah Rizg (2022), Smart Agriculture Applications Using Deep Learning Technologies: A Survey; *Appl. Sci.*, 12(12), 5919; <https://doi.org/10.3390/app12125919>.
7. Marwan Albahar (2023), A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities; *Agriculture* 13, 540. <https://doi.org/10.3390/riculture13030540>.
8. Rosana Cavalcante de Oliveira & Rogério Diogne de Souza e Silva (2023), Artificial Intelligence in Agriculture: Benefits, Challenges, and Trends; *Appl. Sci.* 2023, 13, 7405, <https://doi.org/10.3390/app13137405>.
9. R. Krishnamoorthy, R. Thiagarajan, S. Padmapriya, I. Mohan, S. Arun & T. Dineshkumar (2023), Applications of Machine Learning and Deep Learning in Smart Agriculture; **Page(s): 371 – 395 DOI: [10.1002/9781119861850.ch21](https://doi.org/10.1002/9781119861850.ch21)** (Wiley-IEEE Press)
10. Vishal Meshram, Kailas Patil, Vidula Meshram, Dinesh Hanchate & S.D. Ramkteke (2021), Machine learning in agriculture domain: A state-of-art survey; *Artificial Intelligence in the Life Sciences* (Elsevier publication)

## A Study on Covid -19 Classification From X-Ray Images With Machine Learning

Siddharth K. Ganvir <sup>[1]</sup> and Dr.G. K. Reddy <sup>[2]</sup>

Department of Electronics, Mahatma Fule Arts Commerce and Sitaramji Choudhary Science

Mahavidyalaya, Warud, Dist.-Amravati (MS)

E-mail - [siddharthgnvr@gmail.com](mailto:siddharthgnvr@gmail.com) , [reddygk\\_2007@rediffmail.com](mailto:reddygk_2007@rediffmail.com)

### ABSTRACT:

The emergence of novel variants of the SARS-CoV-2 virus, such as Omicron and Delta, has heightened the urgency for efficient and accurate diagnostic tools. This research proposes a novel approach for automated detection of COVID-19 variants using X-ray images and deep learning techniques. The study leverages a comprehensive dataset encompassing X-ray images from patients infected with different SARS-CoV-2 variants, with a particular focus on the highly transmissible Omicron and Delta variants.

The deep learning model architecture is designed to capture unique radiological features associated with each variant, enabling precise identification. Transfer learning is employed with state-of-the-art pre-trained models to leverage learned representations from diverse datasets. The model's training process is meticulously optimized, considering the challenges posed by variant-specific imaging patterns.

The proposed system undergoes rigorous evaluation on a diverse and well-annotated dataset, encompassing cases of Omicron, Delta, and other prevalent strains. Performance metrics including accuracy, sensitivity, specificity, and F1 score are used to assess the model's effectiveness in distinguishing between different variants. Furthermore, the study delves into the interpretability of the model's decisions, employing attention mechanisms to highlight the regions within X-ray images that contribute significantly to variant classification. This not only enhances diagnostic transparency but also aids in understanding the radiological nuances associated with distinct variants.

The findings hold promise for timely and accurate identification of emerging variants, thereby facilitating targeted public health interventions and clinical management strategies. Additionally, the model's interpretability enhances its acceptability among healthcare professionals, fostering trust in its clinical applications.

**Keywords:** MATLAB, Deep learning model , X-ray imaging , COVID-19, Omicron & Delta variants etc.

### INTRODUCTION:

The ongoing battle against the COVID-19 pandemic has been marked by the evolution of the SARS-CoV-2 virus, giving rise to distinct variants with unique genomic signatures. Among these, the Omicron and Delta variants have emerged as significant challenges due to their increased transmissibility and potential for immune escape. Rapid and accurate identification of these variants is crucial for effective public health responses and clinical management.

Traditional diagnostic methods often face limitations in discerning between different viral strains promptly. This research focuses on advancing the state-of-the-art in COVID-19 variant detection by harnessing the power of X-ray imaging and deep learning, implemented using MATLAB as the development environment. The utilization of neural networks, a subset of deep learning models, offers a promising avenue for automating the identification of

Omicron and Delta variants based on radiological patterns captured in X-ray images. MATLAB, with its robust set of tools and frameworks for image processing and deep learning, provides an ideal platform for developing and implementing sophisticated algorithms. The synergy between MATLAB and neural networks creates a dynamic computational environment conducive to training models on large and diverse datasets, capturing intricate patterns associated with each variant.

This research explores the development of a neural network-based model that leverages MATLAB's capabilities for image preprocessing, feature extraction, and model training. By incorporating a diverse dataset encompassing X-ray images from patients infected with different variants, including Omicron and Delta, the model is tailored to discern variant-specific radiological patterns.

The contributions of this study extend beyond algorithmic development. The MATLAB-based implementation allows for an accessible and replicable framework, potentially facilitating widespread adoption of the proposed methodology. The integration of neural networks trained on X-ray images presents a non-invasive and efficient approach to variant detection, overcoming some of the limitations of traditional diagnostic methods.

In summary, this research endeavors to exploit the synergy between X-ray imaging, deep learning, MATLAB, and neural networks for the automated detection of Omicron and Delta variants of the SARS-CoV-2 virus. The subsequent sections will delve into the methodology, dataset, model architecture, and experimental results, shedding light on the potential of this integrated approach in enhancing our capabilities for variant-specific identification in the ongoing fight against the COVID-19 pandemic.

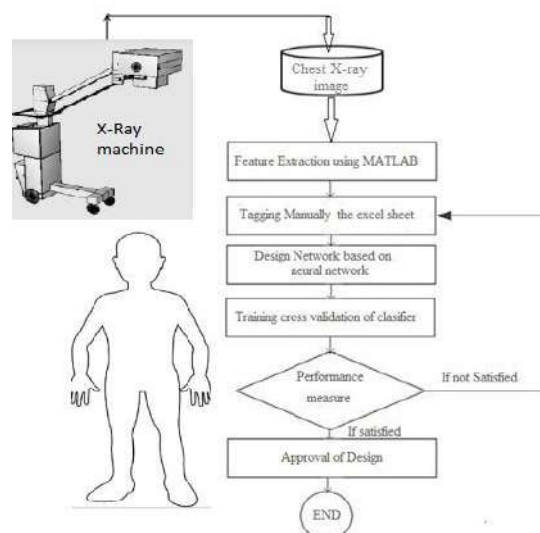
## V. RELATED WORK:

Nurbaiti Sabri , Raseeda Hamzah , Shafaf Ibrahim & Khyrina Airin Fariza Abu Samah [1] has analyzed a total of 101 data consist of 33 COVID-19, 28 normal and 40 bacteria of chest X-Ray images are tested. The extracted features are tested using Weka software where it able to analyze the accuracy of k-NN classifier using 10-folds cross-validation. The result represents with true positive (TP) and false positive (FP) for all tested images. In their research, 3776 attributes from x-ray images have been used for classification purpose. Matrices used to measure the efficiency of the classifier precision and recall. Precision represent the percentage of x-ray images that are classified as true. Meanwhile, recall is the percentage of relevant x-ray images that labeled as "true" by the classifier Result of classification.based. They have clearly observed that LBP able to produce a good classification accuracy with average of 0.960. 96%. The analysis also reveals that the maximum precision and recall are obtained for the LBP algorithm, with both values at 0.96. Sadman Sakib, Tahrat Tazrin, Mostafa M. Fouda, Zubair Md. Fadlullah And Mohsen Guizani [3] proposed DL-CRC framework consists of two parts: the DARI algorithm (which adaptively employs a customized generative adversarial network and generic data augmentation techniques such as zoom and rotation) and a two-dimensional convolutional neural network (CNN) model. They employed a unique dataset for multiple publicly available sources, containing radiograph images of COVID-19 and pneumonia infected lungs, along with normal lung imaging. The classification accuracy significantly increased to 94.61% by adopting our proposed DL-CRC framework. They have compared their proposal with existing deep learning models from di-verse categories such as depth-based CNN (e.g., Inception-ResNet v2), multi-path-based CNN (DenseNet), and hybrid CNN (ResNet) architectures. Extensive experimental results demonstrated that their proposed combination of DARI and custom CNN-based DL-CRC framework significantly out-performed the existing architectures. Thus, incorporating the proposed model with significantly high accuracy into the clinical-grade as well as portable X-ray equipment can allow an automated and accurate detection of COVID-19 in the scrutinized patients.



## PROPOSED METHODOLOGY:

It is proposed to think about the grouping of three dataset images Using Neural Network Approaches.. Information securing for the proposed classifier intended for the Recognition of three type of dataset images .The most vital un corresponded includes and in addition coefficient from the images will be extricated . In order to extract features, statistical techniques, transformed domain will be used.



**Fig 1:** Flow Chart

Computational Intelligence techniques include the following will established techniques.

- i) Statistics
- ii) Learning Machines such as neural network .
- iii) Transformed domain techniques such as FFT, WHT, HISTOGRAM etc.

For choice of suitable classifier following configuration will be investigated.

- i) Support Vector Machine.
- ii) Modular Neural network.
- iii) Generalized Feed Forward Neural Network

For each of the architecture, following parameters are verified until the best performance is obtained.

- i) Train-CV-Test data
- ii) Variable split ratios
- iii) Retraining at least five times with different random initialization of the connection weights in every training run.
- iv) Possibility different learning algorithms such as Standard Back-Propagation, Conjugate gradient algorithm , Quick propagation algorithm, Delta Bar Delta algorithm, Momentum
- v) Number of hidden layers
- vi) Number of processing elements of neurons in each hidden layer.

After regions training & retraining of the classifier, it is cross validated & tested on the basis of the following performance matrix such as Mean Square Error, Normalized Mean Square Error, Classification accuracy, Sensitivity & Specificity

## CONCLUSION:

In the pursuit of advancing automated detection methodologies for distinct variants of the SARS-CoV-2 virus, this research has successfully explored the integration of X-ray imaging, deep learning, MATLAB, and neural networks. The focus on the Omicron and Delta variants, which have posed significant challenges to global health systems, underscores the urgency for robust and efficient diagnostic tools. The developed neural network-based model, implemented in MATLAB, demonstrated promising results in variant-specific identification. Leveraging the inherent capabilities of MATLAB for image processing, feature extraction, and deep learning, the model exhibited a commendable ability to discern subtle radiological patterns associated with different SARS-CoV-2 variants. This success paves the way for non-invasive and timely identification, contributing to the arsenal of tools available for frontline healthcare professionals.

The interpretability of the model was a critical aspect of this research. By incorporating attention mechanisms, the study sought to enhance transparency in the decision-making process, providing insights into the regions of X-ray images pivotal to variant classification. This not only contributes to the trustworthiness of the model in clinical applications but also facilitates a deeper understanding of the radiological nuances linked to specific viral variants.

Furthermore, the MATLAB-based implementation offers a practical and accessible framework for future research and potential clinical deployment. The adaptability and scalability of the proposed methodology can potentially contribute to the development of diagnostic tools that are both effective and widely applicable. While this research presents a significant step forward, it is essential to acknowledge its limitations. The dataset used, although comprehensive, may benefit from continuous expansion to encompass a broader spectrum of cases and demographics. Additionally, ongoing surveillance for emerging variants necessitates the model's adaptability to evolving viral landscapes.

In conclusion, the synergistic integration of X-ray imaging, deep learning, MATLAB, and neural networks holds substantial promise in the automated detection of SARS-CoV-2 variants. The outcomes of this study contribute to the ongoing global efforts in combating the COVID-19 pandemic, providing a foundation for further research, refinement, and potentially, real-world applications. As we continue to navigate the complexities of viral evolution, this research stands as a testament to the potential of interdisciplinary approaches in addressing public health challenges.

## VI. REFERENCES:

- [1] Nurbaity Sabri , Raseeda Hamzah , Shafaf Ibrahim & Khyrina Airin Fariza Abu Samah (2020), :COVID-19 Detection for Chest X-Ray Images using Local Binary Pattern International Journal of Emerging Trends in Engineering Research, volume 8.No.1.1.
- [2] A. Mangal, S. Kalia, H. Rajgopal, K. Rangarajan, V. Namboodiri, S. Banerjee and C. Arora., (2020), CovidAID: COVID-19 detection using chest x-ray, arXiv preprint arXiv:2004.09803.
- [3] Sadman Sakib, Tahrat Tazrin, Mostafa M. Fouda, Zubair Md. Fadlullah And Mohsen Guizani , DL-CRC: Deep Learning-based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach, DOI 10.1109/ACCESS.2020.3025010, IEEE Access
- [4] C. Qin, D. Yao, Y. Shi and Z. Song, (2018), Computer-aided detection in chest radiography based on artificial intelligence: a survey, Biomedical Engineering Online, vol. 17, pp. 113,.
- [5] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz, and B.N. Patel, (2018), Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, PLoS medicine, vol. 15,.

## AGNIRAKSHAK: Raspberry Pi based fully Robust fire Sensing & Protection System

Siddharth K. Ganvir <sup>[1]</sup>, Dr.S.P.Deshpande <sup>[2]</sup> & Dr.G. K. Reddy <sup>[3]</sup>

Shree H.V.P.Mandal's , Degree College of Physical Education, Amravati(Maharashtra)

E-mail – siddharthgnvr@gmail.com , shrinivasdeshpande68@gmail.com ,reddygk\_2007@rediffmail.com

### ABSTRACT

Sensing and controlling system is having very vast applications in various fields. It used advanced technology such as embedded or Microcontroller based instruments. Now-a days the cases of fire catching have been increased a lot. The places like Colleges, Hospitals, Marriage halls, go-downs are usually got prone to such fire incidences as per statistics. AGNIRAKSHAK is a model for fire sensing as well as to control that fire through a completely robust electronic system. This is able to detect the fire & take the preventive actions accordingly. The Raspberry Pi based model consist of a Flame sensor, Smoke Sensor ,Temperature Sensor, Alert Sound Buzzers, Automatic sprinkler (water pour) system, GPS with GSM module for sending location and a help message to nearby fire brigade agencies and a physical rescuing outlets system. This "Fully robust fire detection & protection electronic system" has been designed in such a manner referencing recent fire catch incidences in mind. When the flame sensors sense the flame the Raspberry Pi module triggers the connected circuits due to which power within the place gets cutoff, Buzzers starts alerting about fire, water sprinklers gets activated, help message with location will get send to nearby fire brigade agencies. Ladders comes in ready to use position fixed at the windows.

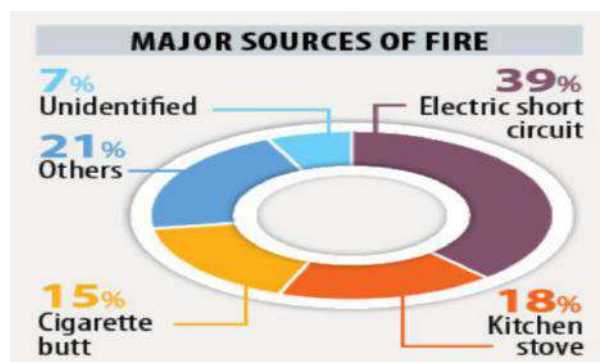
Keywords: Raspberry Pi, Sensors, Fire etc

### INTRODUCTION

The risk of fire in urban areas has increased over the years and the rise in cost of fire losses are increasing at a greater rate than the measures devised to control them. Human negligence also plays a key role in such accidents. The proper laws should be in place to tackle the issue. As the country is facing many hazards like biological, chemical, knowing about fire incidents and fire safety is important.

### DESCRIPTION OF PROBLEM

The use of fire are everywhere in the world , Sometimes due to some natural as well as man-made error the place meet with the fire accidents. The sources of fires are expressed approximately in terms of percentages.



The existing fire safety system that are already in use are not enough to meet the requirement. It has been observed from the past fire accidents that the existing system needs to be modified with new technology making use of electronics communication. Also the physical traditional rescue ladders are not of enough length & the fire brigade takes too much time to reach the accidental place.

The Building industry is undergoing transformation with the advancements in sensors, electronics, information communication & technologies. As a result many new technologies have emerged, Similarly, with the advancements in material & insulation technologies & their extensive use in building, the risk to life & property is also increasing due to fire. Following figure shows the statistics of fires in India ADSI Report by NCRB,2020.

Thus there is increased need of the development of Intelligent buildings and systems enabled with fire detection as well as protection systems.

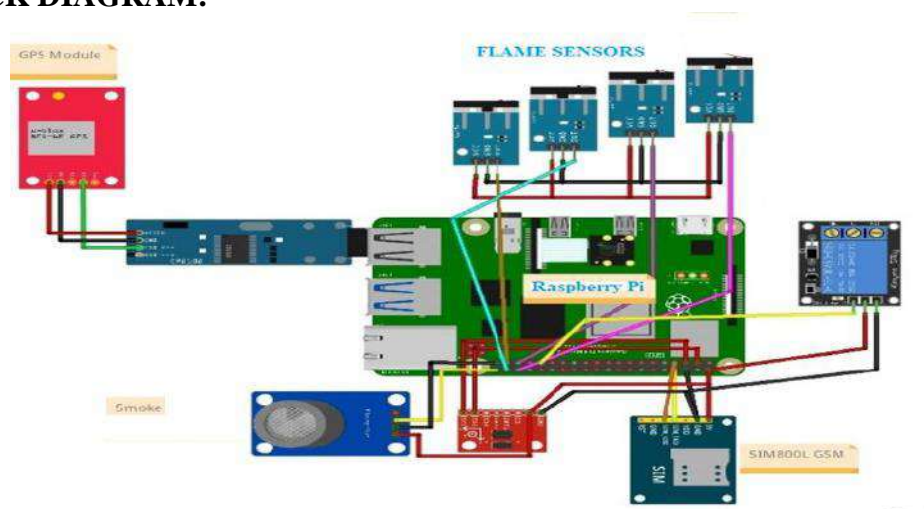
<b>FIRES IN INDIA</b>				
	<b>Fire accidents</b>		<b>Deaths</b>	<b>Injuries</b>
2016	16,695		16,900	998
2017	13,397		13,159	348
2018	13,099		12,748	777
2019	11,037		10,915	441
2020	9,329		9,110	468

Source: ADSI report by NCRB, 2020

## OBJECTIVES

- To detect the fire & to take real time preventive actions accordingly.
- Minimum Time delay in getting out from the fire place if trapped specially at Building having multiple floors.
- To reduce the losses of life & wealth at the public due to fire.
- To increase the fire safety at the public places such as Hospitals, Schools, Colleges, Coaching Institutes, Cinema Theatres, Shopping Malls, Hotels, Temples, Marriage Halls, Factories etc.

## BLOCK DIAGRAM:

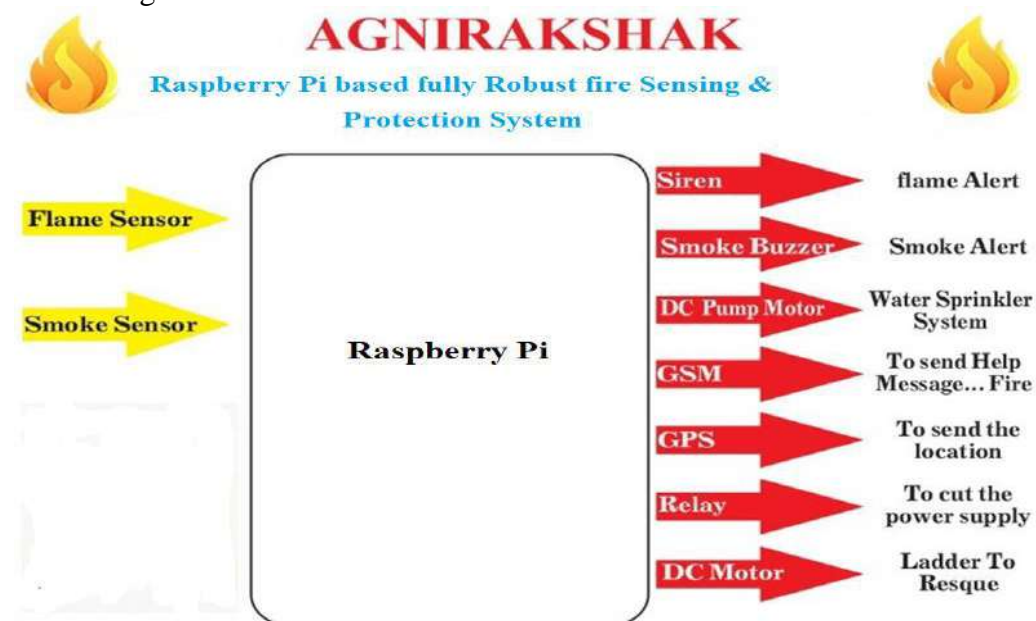


## COMPONENTS USED

- **Flame Sensor-** The Flame Sensor Will going to sense the Flame
- **Raspberry Pi** –It is a series of small single-board computers (SBCs) developed in the United Kingdom by the Raspberry Pi Foundation in association with Broadcom. Since 2013, Raspberry Pi devices have been developed and supported by a subsidiary of the Raspberry Pi Foundation, now named Raspberry Pi Ltd. The Raspberry Pi project originally leaned toward the promotion of teaching basic computer science in schools
- **Smoke Sensor-** The Smoke Sensor Will going to sense the Smoke
- **Raspberry Pi-** It has a dedicated processor, memory, and a graphics driver, just like a PC. It also comes with its operating system, Raspberry Pi OS, a modified version of Linux.
- **GSM Module-** It is used to send the HELP message to nearby fire brigade agencies
- **GPS Module-** It is used to send the location information.
- **Relay System-** Relay is a electrically operated switch.which is going to switch of the power Supply.Also will make the water sprinkler system on through DC pump motor & will also push The ladder to downward direction in ready to use position.

## WORKING

The Raspberry pi based model consist of a Flame sensor, Smoke Sensor, Alert Sound Buzzers, Automatic sprinkler (water pour) system, GPS with GSM module for sending location and a help message to nearby fire brigade agencies and a physical rescuing outlets system. This “Fully robust fire detection & protection electronic system” has been designed in such a manner referencing recent fire catch incidences in mind.



When the flame sensors sense the flame the Raspberry pi triggers the connected circuits due to which power within the place gets cutoff, Buzzers starts alerting about fire, water sprinklers gets activated, help message with location will get send to nearby fire brigade agencies. Ladders comes in ready to use position fixed at the windows.

---

## CONCLUSION

The Raspberry pi based fully robust fire sensing and protection system will be a finest solution for any place where there are the often exposure of fire through any means. The robustness of the proposed model will surely contribute not only to the safety of mankind but also in saving the lives. Also this system will generate the real time 'alert & help' signals to nearby fire brigade agencies.

It is highly suitable for all the public places such as Hospitals, Schools, Colleges, Coaching Institutes, Cinema Theatres, Shopping Malls, Hotels, Temples, Marriage Halls, Factories etc. The damage due to fire involving damages of wealth, infrastructure, and human lives can be minimized by incorporating the Agnirakshak at most of the public places

## REFERENCES

1. Brushlinsky P.W.N., Ahrens M., Sokolov S. World Fire Statistics, (2019). [(accessed on 15 May 2021)]. Available online: <https://www.ctif.org/>
2. Joo J.-Y., Ilić M.D, (2016) , An information exchange framework utilizing smart buildings for efficient microgrid operation. *Proc.IEEE* 104:858864.doi: 10.1109/JPROC.2016.2526119. [[CrossRef](#)] [[Google Scholar](#)]
3. Morgan A. New Fire Detection Concepts. *Fire Saf. Eng.* 2000;**7**:35–37. [[Google Scholar](#)]
4. Liu Z., Makar J., Kim A.K., Development of fire detection systems in the intelligent building. *NIST Spec. Publ. SP.* 2001;2001:561–573. [[Google Scholar](#)]
5. Crapo W.R. Smoke Detectors and Life Safety. *Fire Eng.* 2002;153:61–69.

## Ladder Logic Diagram – A PLC Programming Method for Automation

**Mr. A. G. Kshirsagar**

Research Scholar  
Department of Electronics  
Brijlal Biyani Science College  
Amravati

**Dr. G. D. Agrahari**

Professor and Head  
Department of Electronics  
Brijlal Biyani Science College  
Amravati

**Dr. D. S. Dhote**

Professor and Principal  
Brijlal Biyani Science college  
Amravati

**Mr. M. C. Naidu**

Assistant Professor, Department of Electronics, Hislop College, Nagpur  
E-mail – kshirsagar11a@gmail.com

### Abstract

A unique graphical technique used to design a specific control system or process is known as a ladder logic diagram. The paper covers the fundamentals of ladder programming which includes various symbols, rules of drawing, understanding the concept of ladder logic, various terminologies used with programming, etc. Using ladder logic diagram, it is possible to convert an electrical control circuit diagram to a PLC program and saves the number of various physical components. The various ladder logic programming software is available to program a PLC but the selection of particular software is dependent on PLC selection. Identification of all inputs and outputs, conditions of process or outputs controlling are the key points in the designing of ladder logic diagram. Ladder logic is used to develop a program for PLC to execute the process or operation sequentially. There are various programming languages available to program a PLC but ladder logic is most commonly used because of the number of advantages like simplicity to understand, flexibility, reliability, easy troubleshooting, easy graphical representation, etc.

**Keywords** – Automation, Ladder logic diagram, Rules of ladder programming, System design and terminologies concern with ladder programming.

### Introduction

Automation is an approach designed to control processes with less human assistance or without any assistance from humans. Various sensors, switches, actuators, computers, etc replaced human assistance or made it less. With the help of technology, automation makes the process easier. Consistency in process, quality-based products, less cost, increase in productivity, error-free process, and time management are some of the aspects behind the implementation of industrial automation. Nowadays automation has become a part of every industry. In industrial automation, the various industrial equipment and machines are controlled with the help of logical programming. Ladder logic programming is the most commonly used language for PLC programming. The construction of a ladder logic program includes some basic parts which include rails, rungs, inputs, outputs, address notation, logical expressions, etc. To understand the ladder logic, it is important to know that the flow of execution starts from left to right in a rung, and process execution starts from top to bottom in a rail.

### Rules for Ladder Logic Program Designing

- (1) Inputs can be connected in series as well as parallel.
- (2) Outputs are connected only in parallel
- (3) The input address never be used as an output address.

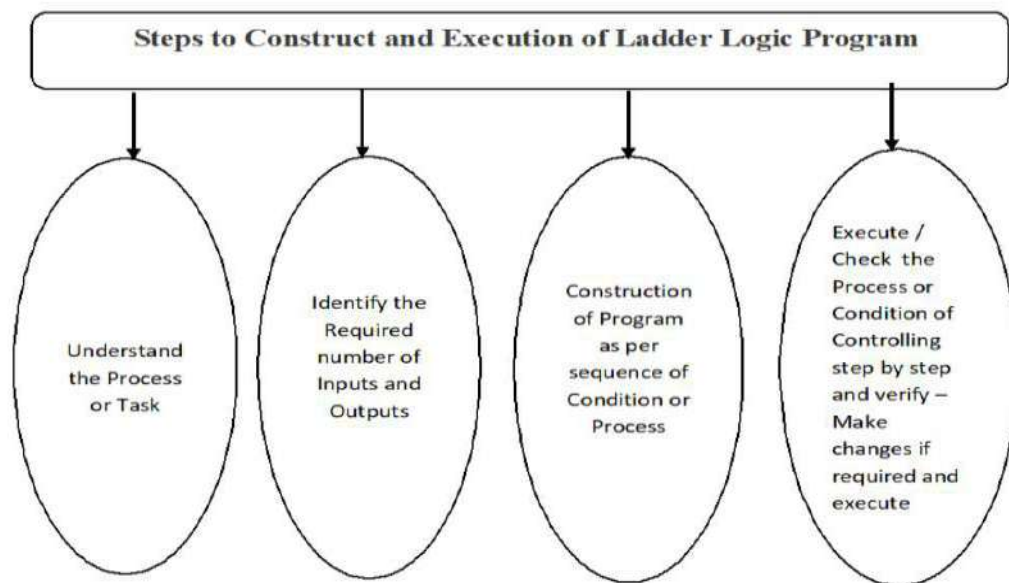
- (4) The output address can be used as an input address.  
 (5) One input can be used several times.  
 (6) One output cannot be used several times (except Set or Reset and Latch or Unlatch)

#### Ladder Logic diagram Symbols

Symbol	Function	Symbol	Function
	Normally Open Contact		Down Counter
	Normally Closed Contact		Up Counter
	Output Coil		Adder
	Set Coil		Subtractor
	Reset Coil		Division
	One Shot / Positive edge		Multiplication
	Timer Delay ON		PID Controller
	Timer Delay OFF		Greater than or equal to

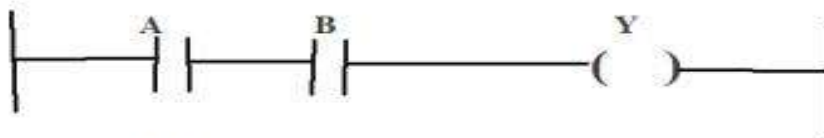
Less than or equal to, Equal to, latch, unlatch, set, reset and many more symbols are used in the ladder programming depending on program logic and PLC.





### Ladder Diagram of Various Logic Functions

#### ❖ AND Ladder Rung



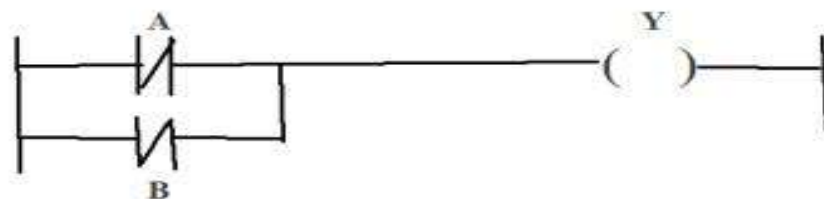
#### ❖ OR Ladder Rung



#### ❖ NOT Ladder Rung



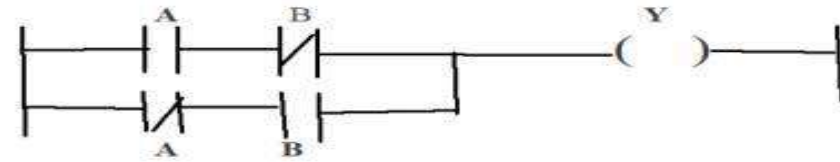
#### ❖ NAND Ladder Rung



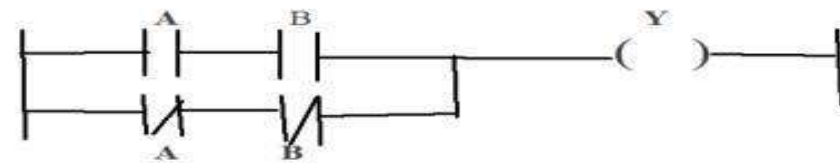
❖ **NOR Ladder Rung**



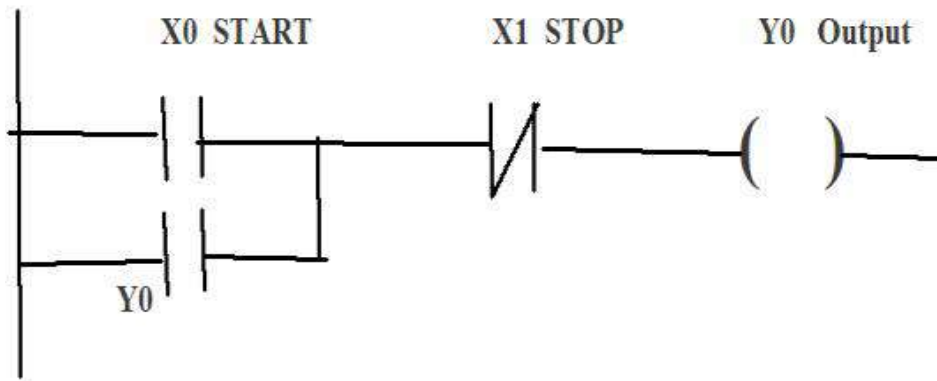
❖ **EX OR Ladder Rung**



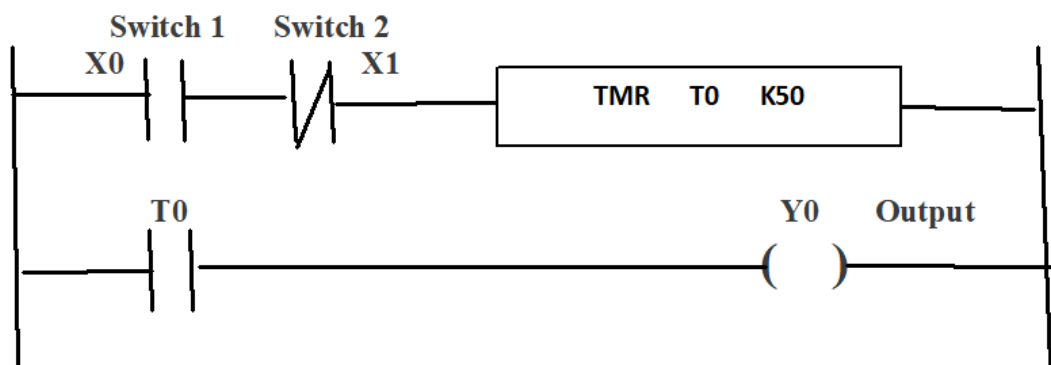
❖ **EX NOR Ladder Rung**



**START and STOP Logic for Push Button (Holding / Latching Concept)**



**Example -When switch 1 is pressed, the timer will be ON and after 5 seconds of timer the output will be ON.**



---

## Outcome

- Using a Ladder logic diagram a simple electrical ladder diagram can be converted into a PLC program.
- Using various symbols of ladder logic programming, the number of physical components can be minimized
- A ladder logic diagram is a graphical representation of a process or task to be controlled that is easy to understand.

## Conclusion

Using a ladder logic diagram, a particular program can be written as per the condition or operation of the process, and controlling of outputs can be done easily. Due to its graphical nature, it is easy to understand as well as troubleshooting or fault finding is easy.

## References

- A. D. Vieira, E. A. P. Santos, M. H. de Queiroz, A. B. leal, A. D. de Paula Neto and J. E. R. Cury, “ *A Method for PLC Implementation of Supervisory Control of Discrete Event Systems* ”, Published by *IEEE Transaction on Control Systems Technology*.
- S, C. Ramesh, Dr. G. Kalivarathan ,January – February (2013),, “ *A Reasonable Approach for Manufacturing System Based On Supervisory Control Using Discrete Event System* ” Published by *IJECET, Volume 4, Issue 1. PP. 92-98*.
- Andre B. Leal, Diogo L.L. da Cruz and Marcelo da S. Hounsell, “ *PLC Based Implementation of Local Modular Supervisory Control for Manufacturing System* ” [www.intechopen.com](http://www.intechopen.com)
- Belen C. Diego, Vidal M. Rodilla, Carlos F. Carames, Anibel C. Moran, Raul A. Santos “ *Applying a Software Framework for Supervisory Control of a PLC – based Flexible Manufacturing Systems* ” Published Online: 27 October 2009, Springer – Verlag London Limited 2009, *Int J adv Manuf Technol* ( 2010).
- Y. Qamsane, M. El. Hamlaoui, A. Tajer, and A. Philippot (2017) “ *A Model–Based Transformation Method to Design PLC – Based Control of Discrete Automated Manufacturing Systems* ” *Proceedings of Engineering and Technology – PET, Vol. 19, pp. 4 – 11, ISSN 2356 – 5608, 4<sup>th</sup> International Conference on ACECS – 2017*.
- X. D. Koutsoukos, P. J. Antsaklis, J. A. Stiver and M. D. Lemmon, (2000) “ *Supervisory Control of Hybrid Systems* ”, *Proceedings of the IEEE, VOL. 88, NO. 7, July 2000*.

## 11

**Third Eye : Arduino Based Aid for Blind****Mr. Nilesh R. Bundile<sup>1</sup>**

Department of Electronics,  
Vidya Bharati Mahavidyalaya,  
Amravati  
[bundilenilesh66@gmail.com](mailto:bundilenilesh66@gmail.com)

**Mr. Chandrakant R. Chaudhari<sup>2</sup>**

Research Scholar  
Department of Electronics,  
Mahatma Fule, Arts, Commerce &  
Sitaramji Chaudhari Science  
Mahavidyalaya, Warud  
[itsmechandu04@gmail.com](mailto:itsmechandu04@gmail.com)

**Dr. Rajani J. Gajbe<sup>3</sup>**

Professor & Head  
Department of Electronics,  
Vidya Bharati  
Mahavidyalaya, Amravati  
[rjgajbe20@gmail.com](mailto:rjgajbe20@gmail.com)

**ABSTRACT:**

One of the biggest concerns of people who are blind or partially blind is detecting obstacle. The World Health Organization (WHO) estimates that 200 billion people have low vision and approximately 30 million people are completely blind. Arduino-based Third Eye, also known as Extra Vision for the Blind, is a hardware and software that uses ultrasonic waves emitted by an ultrasonic sensor to vibrate and identify objects for the user. A buzzer is used to create this vibration. For weak and blind people, this is a wearable device that makes walking easier. You don't need to take anything with you while walking, just wear our products.

**Keywords:** *Ultrasonic Sensor, Aid, Buzzer, Arduino, Blind Stick, Impaired.*

**INTRODUCTION:**

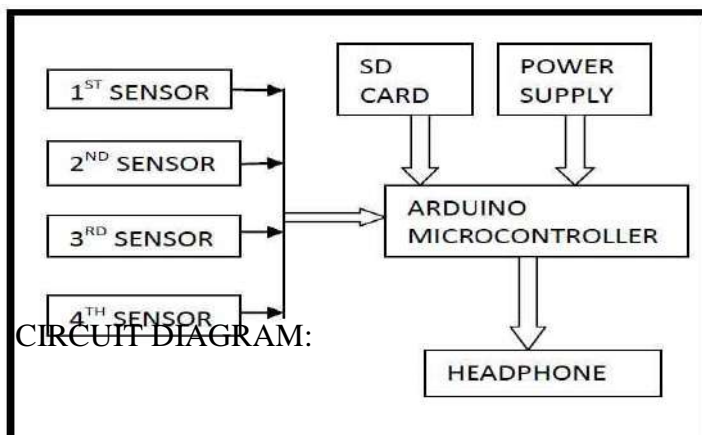
According to the World Health Organization, approximately 180 million people worldwide are blind. They suffer a lot in their daily lives. The affected ones have been using traditional white cane for years, and while it's convenient, it has many drawbacks. Another option is to get a pet such as a dog, but this is very expensive. Therefore, the aim of the project is to create a cheap and effective way to help blind people walk easier, faster and safer. This project aims to create a complete portable device using Arduino that will assist blind individuals and solve problems in existing systems. The system is designed to use ultrasonic sensors to detect objects or problems and emit sound for guidance. The main feature of this technology is that it is the first wearable technology that detects obstacles using ultrasonic waves for blinds and notifies users with a buzzer.

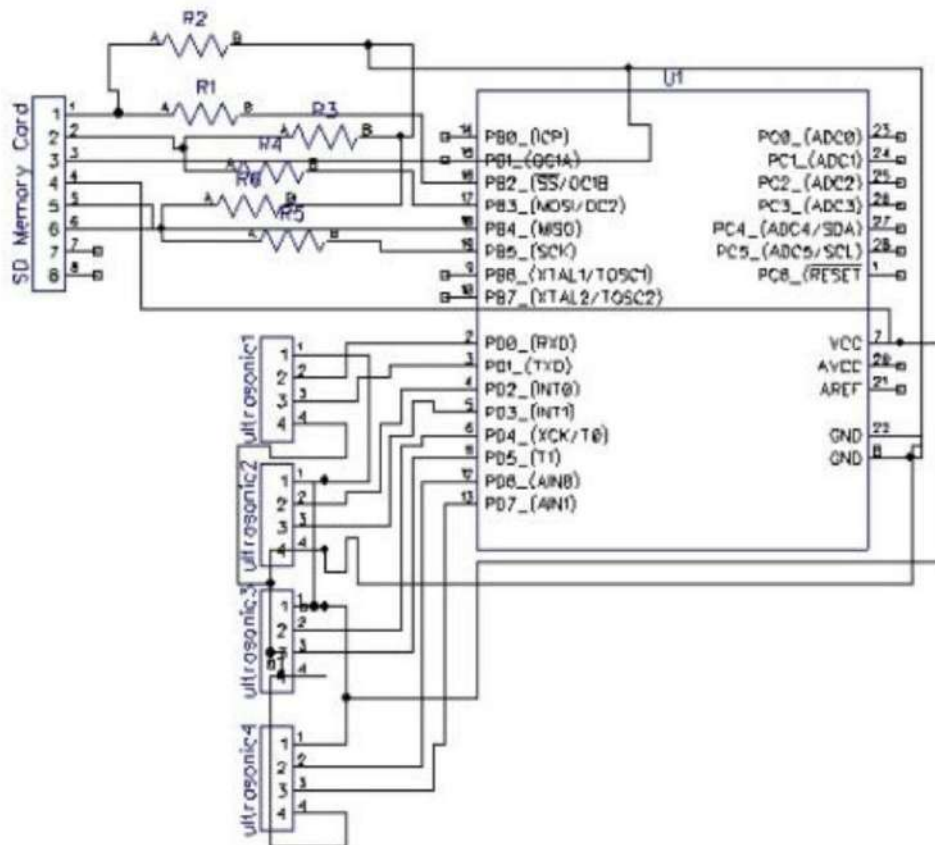
**LITERATURE REVIEW:**

According to the World Health Organization, there are 2.3 billion blind people worldwide, 14% of whom are blind and 86% are visually impaired. One of the most important senses in human life is vision. Vision allows a person to communicate with the outside world. Blind people need additional assistance, such as another person or a cane. Although using memory to get from one place to another is not a safe way for blind people to travel, they can navigate familiar surroundings by keeping maps and potential influences (such as family) for reference. Since most blind people cannot get help from other people, especially if they are outside, a tool such as a stick is needed to help them live in many areas of life. To be effective for blind or partially sighted people, a cane must meet two important criteria: It must be affordable and practical. The purpose of the blind stick is to inform users about the various difficulties associated with the use of the speakers on blind stick. A blind person's path is blocked by obstacles such as people, cars, rocks, as well as obstacles such as stairs, walls, and furniture. When the device detects dampness, it will vibrate and warn the user [1]. By using ultrasonic sensors and buzzers to generate vibrations, blind people can be provided with an Arduino-based third eye (i.e. extra vision), allowing them to identify objects through

correction, hardware and software design. Walking stick: Measure of concern for weak ones: The canes used by visually impaired people for long walks have affected this task. To make walking easier for weak or blind people, our equipment allows them to walk without carrying anything. Arduino's ultrasonic sensor can identify nearby objects and send a signal to the user through a buzzer, allowing them to reach their destination conveniently. Arduino is a software device that includes hardware components such as battery and coding [2]. Modern canes are designed to provide better guidance to blind or visually impaired people. Improving blind stick can give access to the visually impaired by using technology to get around easily [3]. Mixed ultrasonic sensors are connected to the blind stick. The ultrasound concept powers the sensors. When hearing noise occurs, the sensor sends this information to the microcontroller. At this point the microcontroller determines whether the barrier is close enough. If the problem is not near the person, the controller will not show action. Microcontrollers allow the use of buzzing signals when an obstacle approaches the person [4]. Many sensors are also used for the convenience of users. Each sensor collects data and sends it to the Arduino, which generates automatic sounds. Another feature of the system is an embedded module designed to help visually impaired people find their canes. We do this using a wireless remote control equipped with RF. When you touch a special button, the remote gives feedback to the user to help determine the position of the stick, which causes the stick to start buzzing. If the remote is on, the cane will continue to buzz until it turns off. Use GPS modules to help track visually impaired people [5].

#### BLOCK DIAGRAM:





### CONSTRUCTION & WORKING:

The system consists of Arduino Nano, ultrasonic sensor, Perf board, vibration motor, buzzer to check barriers and inform users, red LED, switch, jumper wires, power bank, male and female pins, 3.3 volt old mobile battery of an old unused or discarded phone, some elastic and stickers allow the device to be attached to the user as a band. The connection of the device is as follows. The ground of LED, buzzer and vibration motor is connected to GND of Arduino. The +ve of the LED and the middle leg of the switch are connected to Arduino pin 5. The +ve of the buzzer is connected to the first leg of the switch, and the +ve of the vibration machine is connected to the third leg of the switch. The ultrasonic sensor is wired accordingly. Ultrasonic sensor pin VCC connects to Arduino pin VCC, ultrasonic sensor pin GND connects to Arduino pin GND, ultrasonic sensor pin Trig connects to Arduino pin 12, ultrasonic sensor Pin Echo connects to Arduino pin 12. The switch used for the selection of mode here (ringtone or vibrate mode). First of all, we cut the perforated board in 5 x 3 cm size and solder the Arduino female header to the board. Then connect it to where the buzzer is located. Then place the vibration motor and solder the wires to it. Then connect the LED. Then connect the switch. Connect the header pins to the ultrasonic sensor and battery inputs. Then solder everything and connect the Arduino and ultrasonic sensor to the board. Also add elastic bands to all modules. For the handheld module, use 4 jumper wires to connect the ultrasonic sensor to the circuit board. Then connect the 3.7 volt mobile battery to the board. Then attach the elastic band. Finally, after completing all connections to the Arduino board, upload the code to each Arduino board and use the other 4 modules using a power bank.

The distance between the sensor and the object is calculated as follows:

$$D = (HPTW * SV)/2$$

Where, D = distance in centimetres.

HPTW = high time of pulse width

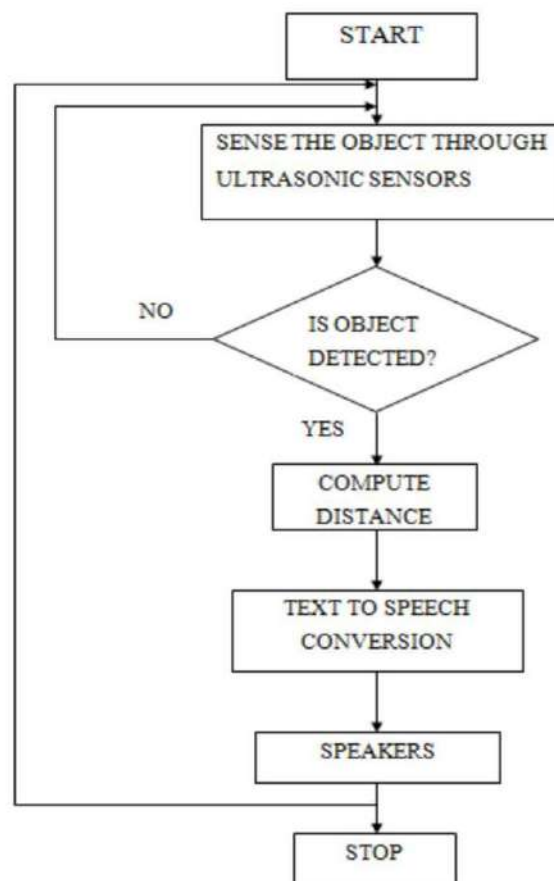
SV = speed of sound in cm/s.

The ultrasonic pulses of the sensor belt must not overlap. Field of view (coverage) measured from a distance of 4 feet is approximately 60 degrees; As the distance from the sensor increases, the relief angle decreases.

So, the aim is to help the blind by scanning the ground with the help of wide-angle sensors and helping them move well.

Therefore, there is an additional process in which the distance is calculated and detected by the sensor and the sound is sent to the user. The working principle of this circuit is shown in flow chart.

### Flow Chart of Third Eye : Arduino Based Aid for Blind



#### APPLICATION:

The World Health Organization (WHO) reports that there are 180 million visually impaired people worldwide. The product is light and compact and will not tire the user. Therefore, this

could help blind people become somewhat self-sufficient by providing information about objects and people in the environment and helping to track their current location.

### **RESULT & DISCUSSION:**

This system is designed and configured for practical use. The system has the ability to control the situations a blind person will encounter. The system will respond to each situation according to a special program coded and installed on the Arduino microcontroller.

It has four features:

1. The obstacle or detector captures the user by sensing vibrations.
2. The moisture detector detects moisture on the ground by beeping quickly so the user can change route.
3. The step or distance detector detects stairs by beeping.
4. The system gives a message when the road is empty.

The ultrasonic sensor on the front can help blind people understand objects when they are close to 9 centimetres. Portability is an important parameter of the system. The system can be worn and used by subject for a long time. The system is easy to use and easy to operate. The system is non-invasive.

### **CONCLUSION:**

This is the first technology for the blind that solves all the problems of existing technology. Nowadays, there are many navigation devices and smart devices for the visually impaired, but most of them have some problems in carrying, and the disadvantage is that they need a lot of training to use. One of the key features of this innovation is that it is affordable for everyone, with its total cost being less than Rs 1,500. There is no such device on the market that is low-cost and convenient to wear like fabric. When a model is used and developed on a large scale, it will benefit society.

### **FUTURE SCOPE:**

Future work will focus on improving physical performance and reducing the burden on users by adding cameras to guide the blind people. Can detect objects/obstacles. It can also determine the material and the form of the object. The single pulse radar principle can be used to identify distant targets. Other possibilities include new ideas based on neural networks that provide better vision and safety for blind people.

### **REFERENCES:**

1. P. Rajesh, R. Sairam, M. D. Kumar, P. K. Eswar and Y. Keerthi, "Arduino based Smart Blind Stick for People with Vision Loss," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1501-1508, doi: 10.1109/ICCMC56507.2023.10083752.
2. Ankush Yadav, Manish Kumar, Vijay Gupta, Shashi Bhushan, "Arduino Based Third Eye for Blind People", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022.
3. Swain, Kunja & Patnaik, Rakesh Kumar & Pal, Suchandra & Rajeswari, Raja & Mishra, Aparna & Dash, Charusmita. (2017). Arduino based automated STICK GUIDE for a visually impaired person. 407-410. 10.1109/ICSTM.2017.8089194.
4. H. Vidhya, A. G. K. Sreeram, K. Kiran and K. Karankumar, "Arduino Based Smart Aiding System for Visually Challenged People," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICPECTS56089.2022.10047229.
5. Harshita Shetty, Richa Khade, Prof. Rohini Bhosale, "DESIGN AND CONSTRUCTION OF ARDUINO BASED AID FOR VISUALLY IMPAIRED PEOPLE", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 07 Issue: 05 | May 2020



## Wireless E-Notice Board using gsm & at mega 328

Anup R. Yenkar<sup>1</sup>, Dr. Rajani J. Gajbe<sup>2</sup>, Chandrakant R. Chaudhari<sup>3</sup>,

Nilesh R. Bundile<sup>4</sup>, Kirti D. Bhanang<sup>5</sup>

<sup>1,2,4,5</sup>Vidya Bharati Mahavidyalaya, Amravati – 444602

<sup>3</sup>Mahatma Fule Mahavidyalaya, Warud (MS)

e-mail : <sup>1</sup>anupyenkar1996@gmail.com

<sup>2</sup>rjgajbe20@gmail.com

<sup>3</sup>itsmechandu04@gmail.com

<sup>4</sup>bundilenilesh66@gmail.com

<sup>5</sup>kirtibhanang14@gmail.com

### ABSTRACT:

The notice board holds significant importance in institutions, organizations, and public spaces such as bus stops, railway stations, or parks. However, the conventional method of sending various notices daily is a cumbersome task. The digital notice board is a hardware solution designed to electronically display notices using a mobile device. Users can input notices between specific characters and transmit them for display, ensuring that only messages within those characters are showcased. This system introduces an SMS-based notice board incorporating the widely used GSM technology to enable message display via the user's mobile phone. Its functionality relies on the ATMEGA328 microcontroller, programmed in assembly language. The microcontroller interfaces with a SIM900 GSM modem, along with a SIM card, utilizing AT commands. When a user sends an SMS from a registered number, it is received by the SIM900 GSM modem on the receiving end. The SIM900 is appropriately interfaced with the microcontroller through a level shifter IC MAX32. The message is then retrieved by the microcontroller and displayed on an electronic notice board featuring an LCD display. The microprocessor is powered by a regulated power supply connected to the mains supply of 230 volts AC.

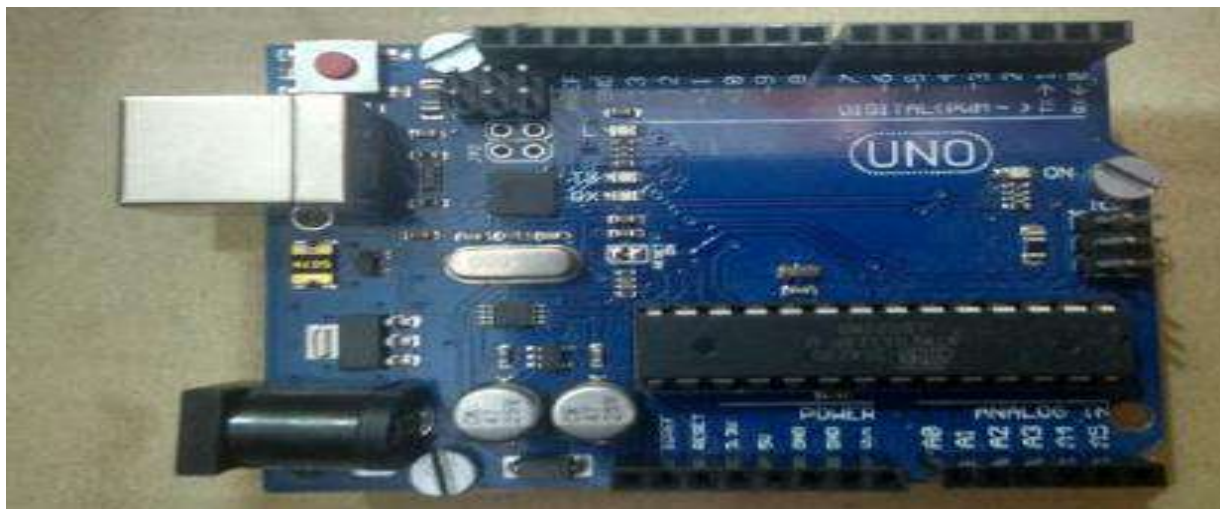
**Keywords:** SIM900 GSM Modem, ATMEGA328, LCD Display

### INTRODUCTION:

A notice board serves as a platform for people to post messages about items for sale, events, or general information. With the widespread use of SMS through mobile phones, there is a growing opportunity to leverage this technology for controlling various devices and displaying data. Using GSM, SMS messages can be received and decoded globally, enabling the control and display of data on LCD monitors from any part of the world. This paper focuses on an advanced wireless notice board, a high-tech solution that utilizes wireless communication. The Wireless E-Notice board project aims to showcase textual messages on a larger display unit with the assistance of a cpp script. When an SMS is sent from a registered number via a mobile phone, the message is received by the SIM900 GSM modem, which is seamlessly interfaced with the Arduino. The Arduino fetches the message, and it is subsequently displayed on an LCD screen connected to the Arduino using an HDMI cable. The messages are programmed to be displayed for a predetermined duration. Additionally, the order of message display is determined by the assigned priority for each sender in the case of multiple senders. The system utilizes an LCD monitor for a more extensive presentation of messages sent via SMS, enhancing the accessibility and reach of the notice board.

### I. SYSTEM OVERVIEW

The block diagram of E-Notice Board consists of-



- 1) Arduino ATMEGA328
- 2) SIM900 GSM Modem
- 3) LCD display

Figure1: Block diagram of Wireless E-Notice Board

- 1) **Arduino ATMEGA 328** : Arduino is an open-source platform utilized for constructing electronics projects. It comprises both a physical programmable circuit board, often termed a microcontroller, and an Integrated Development Environment (IDE) software that operates on your computer. The IDE is employed for writing and uploading computer code to the physical board. Arduino has the capability to interface with buttons, LCDs, motors, speakers, GPS units, cameras, the internet, and even your smartphone or TV. This adaptability, along with the fact that the Arduino software is freely available, the hardware boards are reasonably priced, and both the software

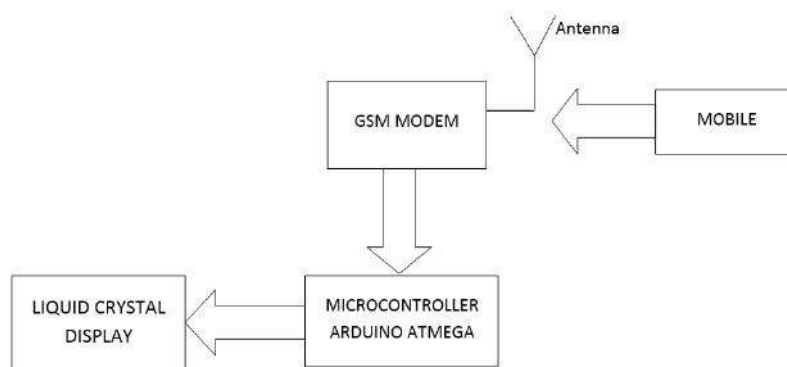


Fig. 2. Arduino UNO

and hardware are easy to grasp, has cultivated a substantial community of users. This community actively contributes code and shares instructions, resulting in a wide array of projects based on Arduino.

**SIM900 GSM Modem** : Global system for mobile communication(GSM)is a wide area wireless communications system that uses digital radio transmission to provide voice, data and multimedia communicating services. A GSM system coordinates the communication between a mobile telephones(mobile stations),base stations(cell sites),and switching systems. Each GSM radio channel is 200KHz wide channels that are further divided into frames that holds 8

time slots. GSM was originally named group special mobile. The GSM system includes mobile telephones (mobile stations), radio towers (base stations), and interconnecting switching systems. The SIM900 is a complete Quad-band GSM/GPRS solution in a SMT module which can be



Fig. 3. GSM Module

embedded in the customer applications. Featuring an industry-standard interface, the SIM900 delivers GSM/GPRS 850/900/1800/1900MHz performance for voice, SMS, Data, and Fax in a small form factor and with low power consumption. With a tiny configuration of 24mm x 24mm x 3 mm, SIM900 can fit almost all the space requirements in your M2M application, especially for slim and compact demand of design.

- 2) **LCD display** : Most LCDs with a controller has 14 pins and LEDs with 2 controller has 16 pin pins (two pins are extra in both for back –light LCD connections) .They have a standard ASCII set of characters and mathematical symbols for an 8-bit data bus, display required a +5v supply plus 11 I/O lines for 4-bit data bus it only requires the supply lines plus seven extra line. When the LCD display is not enable, data lines are tri state and they do not interface with the operation of the microcontroller data can be placed at any location on the LCD.



Fig. 4. LCD Display

## II. Software:

The software used in the proposed project is the Arduino Integrated Development Environment - or Arduino Software (IDE) . It consists a text editor for writing the code, a message area and some other common functions. It connects to the Arduino hardware to upload programs and for communication.

## III. IMPLEMENTATION

Software Implementation :

The software implementation, mainly consists of programming using C language. The AT commands are used to operate the GSM modem. AT command, which is used for authentication consists of a message being preceded by „#“ and followed by „\*“ to indicate start and stop for reading of the message respectively. For example, if the message to be displayed is “Hello”, then it is written as “ #Hello\* ” .

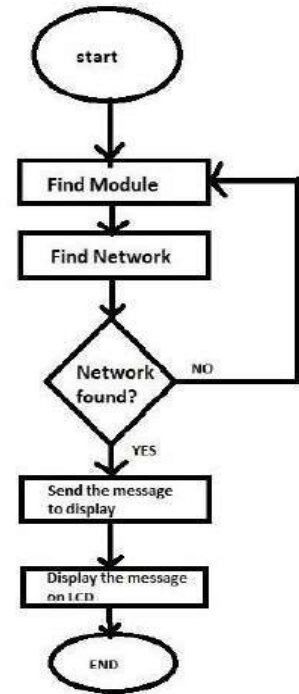
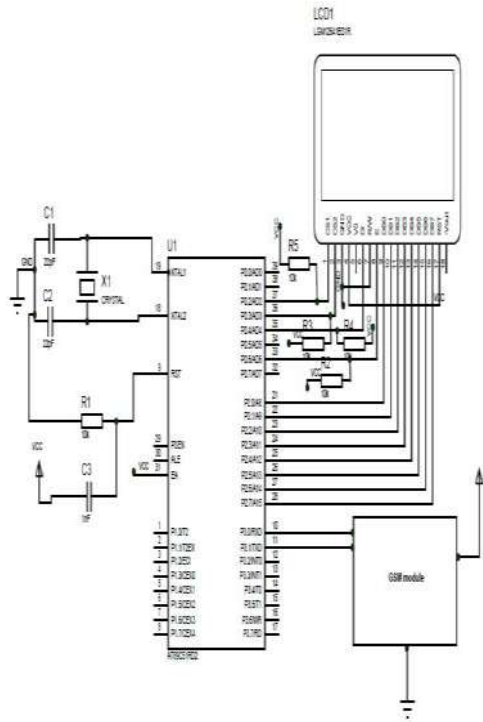


Fig.5 . Flow chart

**CIRCUIT IMPLEMENTATION :**

The connections in the circuit implementation of the proposed system are as shown in theFigure 6. The LCD, which is used for display purposes has its pins namely RS, EN, D4, D5, D6, and D7, connected to the arduino pins. The Arduino pins used for interfacing are 7, 6, 5, 4, 3, and 2 respectively. The Rx and the Tx pins of the GSM module is connected to the Tx and Rx pins of the Arduino respectively for transmission of message. The GSM module is powered using a 12V adaptor.



Fig.6. Circuit design of the model  
RESULTS

When the message send from the mobile, the GSM Modem which is connected to the ATMEGA328PU and the display unit, will receive the message. Now, the ATMEGA328PU reads the message from the GSM Modem and displays it on LCD. When user sends the message from the mobile, GSM modem sends the below command serially to indicate that a new message is received.

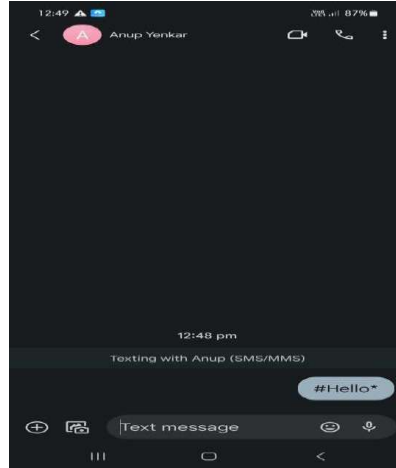


Fig. 8. Message being sent from a mobile device

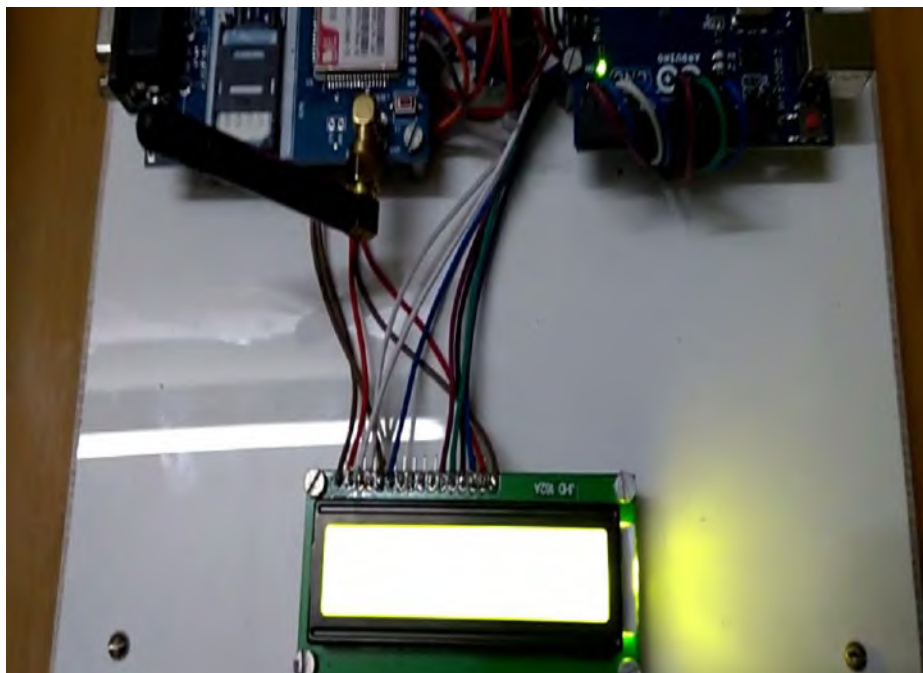


Fig. 9. Message received, displayed on the LCD

### **FUTURE SCOPE**

Temperature display during periods wherein no message is in memory is one improvement that is well possible. Another very innovative and significant improvement would be to accommodate multiple receiver MODEMS at different positions scattered across large geographical areas and carrying duplicate SIM cards. Another added variation in the system can be multilingual display. This feature can be included by programming the microcontroller (here which is Arduino) to use different schemes of encoding and decoding in different areas as per the local language. This will ensure that there is an increase in the number of informed users. We can also consider Graphical display as a long term but Wireless E-Notice Board achievable output. MMS technology along with relatively powerful microcontrollers can be used to carry on the tasks. In system sending messages via GSM network through air interface and displaying it on a LCD by using AT commands. The same principle can be used to control electrical appliances at a distant location. same concept can be used to display the image files or PDF's with the use of better wireless technologies than GSM like Bluetooth or Wi-Fi systems with better extended memories. The above concept of display boards can be used in railway stations, for advertisement in shopping malls, in educational institutes. It can be used for managing traffic in metropolitan cities and other public utility places. In recent years, the LCD has found wide spread use by replacing LEDs because of their declining prices, ability to display numbers, characters & graphics and the ease of programming. The model can be utilized to display temperature in case when there is no message to be display. The message can be first received display in standard language, the same message can be converted to another language and the message can be displayed.

### **CONCLUSION**

The display boards are one of the most important media for transferring information to the maximum number of end users. With the advancement in technology the display board systems are migrating from normal hand written display to digital display. Further to Wireless display units. The concept of this system is to introduce a new technology for notice board display system using GSM technology. A user can send a message from anywhere in the world. This paper deals with development of GSM modem connected wireless notice board system, which displays the desired message of the user through an SMS in most populated or crowded places or remote places. This proposed system has many remarkable applications in educational institutions and organizations, traffic management, crime prevention, railways, advertisements etc. Being user friendly, long range and speedy means of conveying information are major characteristics of this system. By using this proposed ideology, we can improve the security system and also make awareness of the emergency situations and avoid many dangers.

---

## REFERENCES

- 1) Mitesh Santhakumar, Prasad Bhagat,Ujjwal Rajjpurohit, Nitesh Mhatre,Prof VarshaBodade (2016), *Wireless E-Notice Board IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. V (Mar-Apr. 2016), PP 03-07 www.iosrjournals.org*
- 2) Uma Ullas Pradhan, Suma N, Seema Ramachandra and Shilpa S Kulkarni (2017) , *Arduino And Gsm Based Wireless Notice Board ,Department of Electronics, Mount Carmel College ,Carmelight, 13(1):100-113, 2017 p-ISSN 0975-9484 e-ISSN 2395-5538*
- 3) Keshav Kumar, Kumari Ritu, Mrigangna Singh, Mangal V.Patil (2018), *Wireless Display using GSM and Arduino* , *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* © 2018 IJSRCSEIT / *Volume 3 | Issue 5 | ISSN : 2456-3307*
- 4) Fizza Hamid & Nusrat Hamid Shah (2018), *Wireless Notice Board Based On Arduino And Gsm Technology* , *International Journal Of Engineering Sciences & Research technology,ISSN: 22779655[Hamid\*etal.,7(2): February, 2018] Impact Factor: 5.164 IC™ Value: 3.00 CODEN:I*
- 5) Waseem Akhtar, Mohd Amir Umar, Sachin Pandey, Sudhanshu Tripathi, Prashant Ranjan, Pushparaj Singh (2017), *Wireless Electronic LCD Notice Board Using GSM Technology, International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)Volume VI, Issue IV, April 2017 | ISSN 2278-2540*
- 6) Kambale Swapnil, Swami Nilesh, Kadam Punam ,Prof Ghodke Yogesh (2019) *GSM Based Wireless Electronic Notice Board, Vol-5 Issue-2 2019, IJARIIE-ISSN(O)-2395-4396*
- 7) On recent trends in engineering and technology, INDIA

## Water Quality Monitoring Using Atmega 328PU

**Ms. Bhagyashri Bhuyar<sup>1</sup> Mr. Chandrakant R. Chaudhari<sup>2</sup>**

<sup>1</sup>Research Scholar, Department of Applied Electronics, Sant Gadge Baba Amravati University, Amravati

<sup>2</sup>Research Scholar, Department of Electronics, Mahatma Fule, Arts, Commerce & Sitaramji Chaudhari Science Mahavidyalaya, Warud.

[<sup>1</sup>bhagyashribhuyar03@gmail.com](mailto:bhagyashribhuyar03@gmail.com) [<sup>2</sup>itsmechandu04@gmail.com](mailto:itsmechandu04@gmail.com)

### Abstract:

Access to safe drinking water is essential for human health and dignity, playing a crucial role in breaking the cycle of poverty by improving health and enabling productivity. However, global freshwater reservoirs are increasingly threatened by contamination, primarily due to human activities such as open defecation, inadequate wastewater management, and industrial chemical spills. Water quality testing is vital for ensuring access to clean water, as it helps identify safe sources, assess distribution systems, and monitor supplies for safety. Understanding water quality dynamics, influenced by natural variations and human-induced pollutants, is imperative for effective management. Disparities in water management practices are particularly stark in developing countries, where industrial waste and sewage often pollute surface waters without treatment. Addressing these challenges requires coordinated efforts at all levels, including policy reforms, infrastructure improvements, and community engagement. Prioritizing water quality testing and adopting sustainable practices can safeguard this vital resource for future generations.

**Keywords:** Water quality, pH, Mica Analysis, SPLC780C controller.

### Introduction :

Having safe drinking water is a human need and right for every man, woman and child. People need clean water to maintain their health and dignity. Having better water is essential in breaking the cycle of poverty since it improves people's health, strength to work. The quality of our global freshwater supplies is under increased threat of contamination. While water contains natural contaminants, it is becoming more and more polluted by human activities, such as open defecation, inadequate wastewater management, dumping of garbage, poor agricultural practices, and chemical spills at industrial sites. Water quality testing is a tool that can be used to help identify safe drinking water at the source, within a piped distribution system, or within the home. Water testing plays an important role in monitoring the correct operation of water supplies, verifying the safety of drinking water, investigating disease outbreaks, and validating processes and preventative measures.

As water moves through the water cycle, it naturally picks up many things along its path. Water quality will naturally change from place to place, with the seasons, and with the kinds of rocks and soil which it moves through. Water can also be polluted by human activities, such as open defecation, inadequate waste water management, dumping of garbage, poor agricultural practices (e.g., use of fertilizers or pesticides near water sources), and chemical spills at industrial sites.

In developing countries, 75% of all industrial waste and up to 95% of sewage is discharged into surface waters without any treatment.



Even though water may be clear, it does not necessarily mean that it is safe for us to drink. It is important to judge the safety of water by taking the following three types of parameters into consideration:

- Chemical - pH
- Physical - temperature, turbidity

### Literature Survey :

The rapid growth of population has the outcome in the depletion of available means of water and falloff in the water quality. Also the quality of underground water has been infected by weed-killer and fungicides. The rivers in India are getting polluted owing to industrial waste and discharge of untreated sewage. In 2013, Nivit Yadav, "CPCB Real Time Water Quality Monitoring Maintenance". In this method the quality of water in Ganges and Yamuna River is tested by using sensors. Since, they are the most polluted river in our country CPCB plans for analysing the water standards. And this method is more expensive. In 2007, Tuan Le Dinh, Wen Hu, Pavan Sikka, Peter Corke, L. Overs, Stephen Brosman, "Design and Deployment of a Remote Robust Sensor" which gives a brief explanation about the specialties and designing's of sensors. In 2010, Quio Tie-Zhn, briefed the quality monitoring system based on GPRS/GSM module, which collects and sends the data to monitoring centre through GPRS. It is an artificial method collection of data and other process will be done slowly. In 2003, Pavlos Papa Georgiou, "Literature Survey on Wireless Sensor Networks", has analysed the various wireless modes, configurations and networks. He analyses the protocols and layers in Wireless networks. In 2011, Satish Turken, Amruta Kulkarni, "Solar Powered Water Quality Monitoring System using Wireless Sensor Network", The Base station (BS) gathered information from distant remote sensors. The BS associated with ZigBee module was powered by sunlight baseboard (Energy harvesting). In 2015, Liang Hu, Feng Wang, Jin Zhou and Kuo Zhao "A Survey from the Perspective of Evolutionary Process in the Internet of Things", in this article, the new arrival and evolution of the internet is made clear. They recommended using the internet of things and the different techniques were explained. In 2016, M N Barabde, the System is used for determining the physiochemical factors of water quality such as motion, temperature, pH, conductivity, and oxidation lowering potential using ZigBee. In 2016, Pavana N R, Dr. M.C. composed the water quality factors by investigating Wireless sensor networks(WSN) and by using the raspberry Pi module which is used with the Linux version. In 2002, W. Ye, J. Heidemann, D. Estrin, measured about "An Energy Efficient MAC Protocol for Wireless Sensor Networks" which described Connection of sensors and the environmental monitoring applications and it mainly says about the minimum energy consuming in the wireless network applications.

In 2002 J. Hill, D. Culler, "Mica analysed a wireless platform for deeply embedded networks. They clearly explain about the communication of data over In multiple networks and the use of embedded networks without predefined protocols. In 2008, Bergant, A. , Tussling, A.S., Vitkovsky, J.P., Covas, D.I.C., Simpson, A.R., Lambert, M.F analysed the Parameters that affecting the flow and the pressure of water-hammer wave attenuation networks by using the classical theory of water -hammer wave attenuation in different networks by using the classical theory of water. In 2011, Allen, M., Preis, A., Iqbal, M. , Sri rangarajan, S., Lim, H. B. , Girod, L. , Whittle, A.J. "Real-time in- network distribution system monitoring to improve operational efficiency,". It says about the deployment of wireless water sentinel project by using online hydraulic modelling, it ensures the continuous delivery of an essential resource in the various networks. Fig 1, says about the various levels of purity in different years.

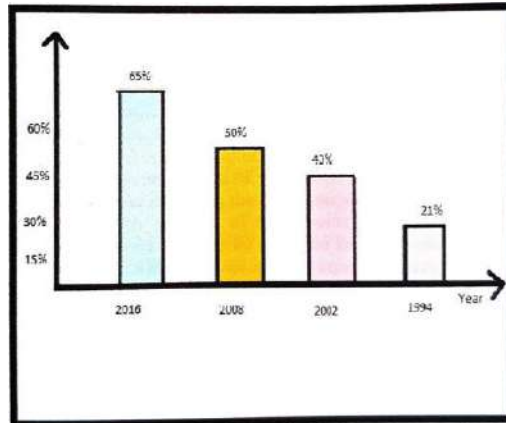


Figure 1 Purity test according to year

### System Design :

The water quality monitoring system employs sensors such as pH, temperature, humidity and turbidity to get the data parameters. These sensors are positioned in the water will analyse the quality of the water resources. The verified content is used to prophesy the quality of water.

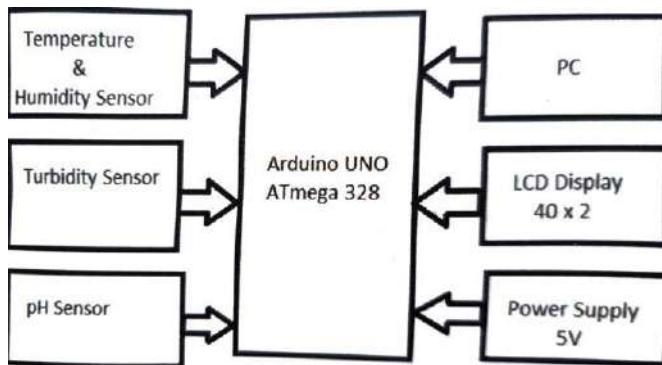


Figure 2 System Design

### Sensors :

A sensor gives a corresponding electrical data by discovering the events or modifications in its environment. A sensor is a transducer device. The Performance of the sensor is increased by the sensor calibration. Speed, accuracy, resolution and linearity are the most important quality of the sensor. The activities can be enhanced & removing of errors due to frame has been deleted in the sensor results which makes it enhance. The difference between the wanted output and the obtained output of the sensor makes way to identify the mistakes due to structure. During the real time measures in the sensor, the repeatable mistakes are compensated during the measured standards.

### pH Sensor:-

It measures the acidic & basic alkaline in the water. It can be defined by using the hydrogen ion concentration with the negative logarithmic. The pH scale range is from 0 to 14, it is logarithmic. The concentration of hydrogen ion values is translated using Ph. The hydrogen ion concentration is small for acidic and if it shows high it is for alkaline solutions. The PH around 7 is the natural source water. The water becomes less acidic as the concentration of hydrogen ion decreases for ten-fold for the increases in the number of PH. A reference electrode & a measuring electrode are enclosed in the pH sensor. The measuring electrode is connected to the positive end of the battery where the reference electrode is connected to the negative

terminal. When the pH sensor is immersed in the solution, the reference electrode has its fixed potential. The change in the hydrogen ion concentration does not change the reference electrode. A potential is developed when hydrogen ion concentration is related to the hydrogen ions which is sensitive to the measuring electrodes. The temperature sensor is necessary to correct any variations in the voltage, as the electrodes differential voltage changes with the temperature.

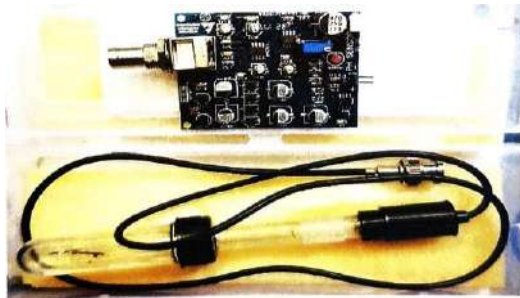


Figure 4 pH Sensor

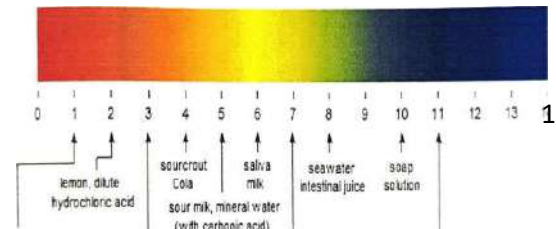


Figure 3 Universal pH Indicator

### Turbidity Sensor:-

Turbidity is the cloudiness or haziness of a fluid which is produced by a large number of independent particles that are generally invisible to the visible eye, like smoke in air. Turbidity is the main method to measure the quality of water. The light that is scattered due to the suspended solids in water is measured by the help of turbidity sensor. When the amount of total suspended solids (TSS) in water increases simultaneously the water's turbidity level (and cloudiness or haziness) also increases. To monitor the turbidity level of water, turbidity sensor is preferred. The gravity Arduino turbidity sensor is preferred to identify the water standard by In measuring the states of turbidity. The sensor uses light to detect suspended particles in water by calibrating the light transmittance and scattering rate and it changes with the quality of total suspended solids (TSS) in water. When the TTS increases by the way the liquid turbidity level also increases. Turbidity sensor is used in measuring the standard of water in rivers and streams, wastewater and the efficient measurements, managing instrumentation for settling ponds, sediment transportation research are also in the laboratory measurements. The analog and digitized signal result modes are given by the liquid sensor. The threshold signal is adjustable when it is in digital signal mode.

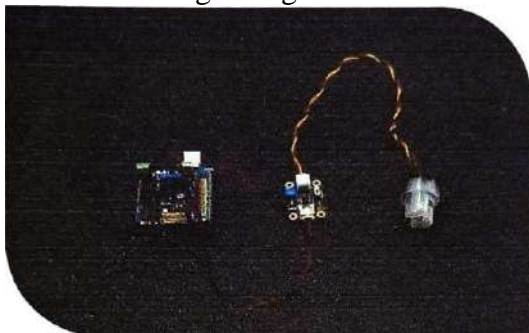
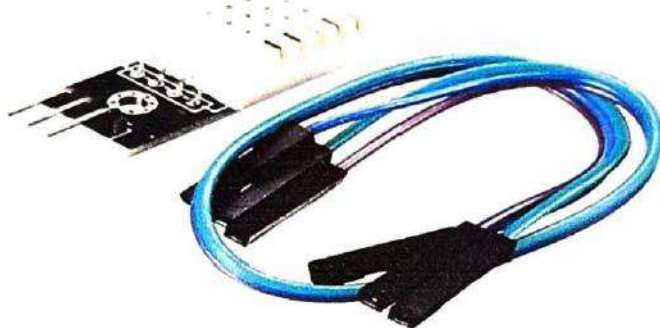


Figure 5 Turbidity Sensor

### Humidity and Temperature Sensor:-

The DHT22 is a basic, low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin (no analog input pins needed). It is fairly simple to use, but requires careful timing to grab data. The only real downside of this sensor is we can only get new data from it once every 2 seconds, so when using our library, sensor readings can be up to 2 seconds old.



*Figure 6 Temperature and Humidity Sensor*

### **Arduino UNO :-**

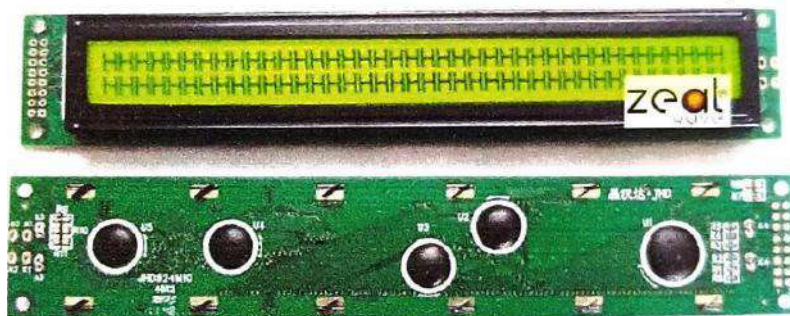
The Uno can be powered via the USB connection or with an external power supply. The power source is selected automatically. External power can come either from an AC-to-DC adapter (wall-wart) or battery. The adapter can be connected by plugging a 2.1 mm centre-positive plug into the board's power jack. Leads from a battery can be inserted in the Ground and Vin pin headers of the POWER connector. The board can operate on an external supply of 6 to 20 volts. If supplied with less than 7V, however, the 5V pin may supply less than five volts and the board may be unstable. If using more than 12V, the voltage regulator may overheat and damage the board. The recommended range is 7 to 12 volts.

### **LCD Screen :-**

IERM4002SYG-1 is 40 characters wide, 2 rows character lcd module. SPLC780C controller (Industry-standard HD44780 compatible controller) has 6800 4/8\_bit parallel interface. A single LED backlight with yellow green colour included can be dimmed easily with a resistor or PWM. Other salient features are as follows:

LCD positive, dark blue text on the yellow green colour, wide operating temperature range, built in character set supports English/Japanese text.

One can refer to the SPLC780C datasheet for the full character set. Furthermore, it has optional 3.3v or 5v power supply and optional pin header connection.



*Figure 7 LCD Screen*

### **1) Test and Measurements :**

Water Quality Monitoring System (WQMS) utilises all three sensors (pH, Turbidity and temperature) and a microcontroller, and finally connecting to the personal computer to process and analyse the quality of water. It is a straight forward and quick device to get the measurement of quality of water.

The project stalled with taking a measurement of each value of pH, turbidity and temperature from each sensor separately. After getting the successful results all the sensors are connected to the microcontroller to make the complete device. Measurement started with a testing temperature of water by taking the hot and cold water with different temperature values. Since the turbidity sensor used in WQMS project is only applicable for qualitative analysis but not the quantitative analysis, only the voltage derived from the sensor is calculated according to the change in voltage with the modification of dirt particles in the water. For measuring the turbidity, dirt particles are added to the water with the random amount and the value of voltage achieved shows as expected. The value of voltage decreases as the dirt particles added to the water which proves that the sensor works as expected. Finally, to measure the pH value, different kind of liquid which contains acid or base is used to test the device.

## 2) Results :-

When the finalised device used to measure the quality of normal tap water, the device provided the pH value of 7.02 and temperature of 23 degree Celsius and providing the turbidity value of close to the pure water (500mV). This is the first test analysed after the device completion. For the confirmation of device working properly, different measurements were taken at the different temperatures, pH value and turbidity. After the device was ready to take the measurements, the parameter used for quality analysis was compared with the reference devices. First of all, the temperature was recorded with the reference thermometer as well as with WQMS at the same water. Cold water was taken at the beginning and slowly little by the little random amount of boiled water (temperature was close to 100<sup>0</sup>C ) was added.

Following Table 1 and Table 2 shows some water quality parameters with respect to various samples of water analysed with the help of our developed system.

Sr. No	Water Sample	Parameters		
		pH	Humidity	Temperature
1	Sample 1 (Tap Water)	8.15	41.90%	34.200C
2	Sample 2 (Rain Water)	9.02	41.90%	34.200C
3	Sample 3 (Bore Well Water)	10.22	41.90%	34.200C
4	Sample 4 (Salt Water)	12.07	41.90%	34.200C
5	Sample 5 (Muddy Water)	9.12	41.90%	34.200C

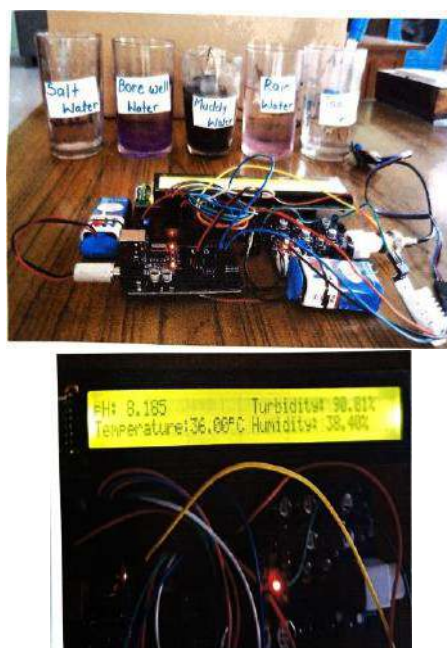


Figure 8 pH Sensor Module and Test Module

Sr. No	Water Sample	Turbidity
1	Sample 1 (Tap Water)	12.79%
2	Sample 2 (Colour Water)	20.54%
3	Sample 1 (Ink Water)	8.34%
4	Sample 1 (Salt Water)	7.05%
5	Sample 1 (Muddy Water)	85.02%

### Conclusion:

Monitoring of Turbidity, PH & Temperature of Water makes use of water detection sensor with unique advantage and existing GSM network. The system can monitor water quality automatically, and it is low in cost and does not require people on duty. So the water quality testing is likely to be more economical-convenient and fast. The system has good flexibility. Only by replacing the corresponding sensors and changing the relevant software programs, this system can be used to monitor other water quality parameters. The operation is simple. The system can be expanded to monitor hydrologic, air pollution, industrial and agricultural production and so on. It has widespread application and extension value. By keeping the embedded devices in the environment for monitoring enables self-protection (i.e., smart environment) to the environment. To implement this, we need to deploy the sensor devices in the environment for collecting the data and analysis.

### Future Scope:

1. In future we can use IOT concept in this project
2. Detecting the more parameters for most secure purpose
3. Increase the water sample parameters by addition of apt sensors
4. By interfacing relay we can control the supply of water.

---

**References:**

1. Caner Borden, Centered Consulting International, LLC, (September 2015), *Dimple Roy International Institute for Sustainable Development (IISD) "Water Quality Monitoring System Design"*, Published by the International Institute for Sustainable Development,
2. A. Qureshi, (2016), *"10 Best Water Quality Testers For Professionals"*, *Wonderful Engineering*,
3. Arduino UNO, ATmega 328, Datasheet
4. Aravinda s. Rao, Stephen Marshall, Jayavardhana Gubbi Marimuthu Palaniswami, Richard Sinnott, and Vincent Pettigrove, (December 2015) *"Design of Low-cost Autonomous Water Quality Monitoring System"*,
5. Reetesh Golhar, Apurva Sakle, Harshali Warhate, Isha Ninawe, Ankita Bodhale, (November 2017) *"The Instantaneous Advanced Water Quality Monitoring System using IOT"*, *International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 11*.
6. Nikhil Kedia, (September 2015) *"Water Quality Monitoring for Rural Areas-A Sensor Cloud Based Economical Project"*, in *1st International Conference on Next Generation Computing Technologies (NGCT-2015) Dehradun, India*,. 978-1-4673-6809-4, 2015 IEEE.
7. Jayti Bhatt, Jignesh Patoliya, (2016) *"IoT Based Water Quality Monitoring IRFIC21"*.
8. Sokratis Kartakis, Weiren Yu, Reza Akhavan, and Julie A. McCann, (2016), *IEEE First International Conference on Internet-of-Things Design and Implementation*, 978-1-4673-9948-7/16 0 2016 IEEE.
9. Mithaila Barabde, Shruti Danve, (June 2015), *"Real Time Water Quality Monitoring System"*, *IJIRCCE*, vol 3, .Eoin O'Connell, Michael Healy, Sinead Thomas and Elie Lewis, *IEEE sensors journal*, vol. 13, no. 7, July 2013, 1530-437, 2013 IEEE
10. Niel Andre cloete, Reza Malekian and Lakshmi Nair, (2016), *"Design of Smart Sensors for Real-Time Water Quality monitoring"*, *IEEE conference*.
11. Pradhnya Bhagde, Mayuri Dabhagde, Aniket Umare, Prof. M. P. Chimurkar, (2018), *"Water Quality Monitoring and Distribution IOT Based Economical Project"*, Vol.4, Issue 6, *IJSRSET*, Page 114- 115.

## Utilization and Generation of Hydropower for Welfare in Agricultural Sector

**Mr. Chandrakant R. Chaudhari<sup>1</sup> Dr. Giridhar K. Reddy<sup>2</sup>**

<sup>1</sup>Research Scholar, Department of Electronics, Mahatma Fule, Arts, Commerce & Sitaramji Chaudhari Science Mahavidyalaya, Warud.

<sup>2</sup>Associate Professor Department of Electronics, Mahatma Fule, Arts, Commerce & Sitaramji Chaudhari Science Mahavidyalaya, Warud.

<sup>1</sup>[itsmechandu04@gmail.com](mailto:itsmechandu04@gmail.com) <sup>2</sup>[reddygiridhar4@gmail.com](mailto:reddygiridhar4@gmail.com)

### ABSTRACT :

This project exemplifies the efficient current hydropower generating at water. Hydropower is an electrical generation technique that harnesses the kinetic energy of flowing water to generate electricity. Since the late 1800s, small-scale hydropower has been a popular method of producing energy in remote areas. Small-scale hydropower systems can be set up in existing water supply networks, such as those for drinking water and wastewater, or in small rivers and streams. Since most small hydropower plants are run-of-river schemes or are integrated into already-existing water infrastructure, they may be erected with little to no environmental impact on species or ecosystems, in contrast to large-scale hydropower systems. In rural or underdeveloped areas, small-scale hydropower is a viable alternative for providing affordable, sustainable energy due to its adaptability, cheap investment costs, and status as a renewable energy source. We are utilising GSM in this project to support wellbeing in the agriculture sector. Global System for Mobile Communication System is abbreviated as GSM. Short message, voice, and data communication are its three primary services. With the aid of this gadget, farmers will be able to set up a wireless communication system, preventing mishaps and harm to the actual model.

*Keywords: Sustainable Energy, Microcontroller 328PU, GSM, Turbidity.*

### INTRODUCTION :

Projects using hydropower have had detrimental effects on biodiversity, the physical environment, and the lives, livelihoods, cultures, and spiritualities of indigenous tribal and illiterate people. Natural resources have been used throughout human history in order to support expanding numbers of people. Increased environmental deterioration and resource exploitation have been factors in the previous several decades in the majority of emerging nations, and if these trends continue unchecked, they may have negative implications on future growth patterns.

Hydropower projects are a renewable energy source that may be used anywhere there is flowing water. They can be small- or large-scale, depending on the local environment and energy requirements. 1.4 billion people worldwide do not have access to electricity, while an additional 1 billion only have sporadic access. Because flowing water is a renewable resource that is present everywhere, using it to create electricity can offer a sustainable energy source that will enhance livelihoods and boost productivity at work. The need to produce more energy from a variety of sources has increased due to the rise in global energy consumption. Currently, hydropower meets around 19% of the world's energy needs.

The creation and application of hydropower is the focus of this project. Small 6 volt generator motor, which produces enough power to run low power applications, is used to create the electricity. The electricity that is not consumed is saved for later use. Energy conservation is in great demand these days since there are so many real-time model applications. Small-scale



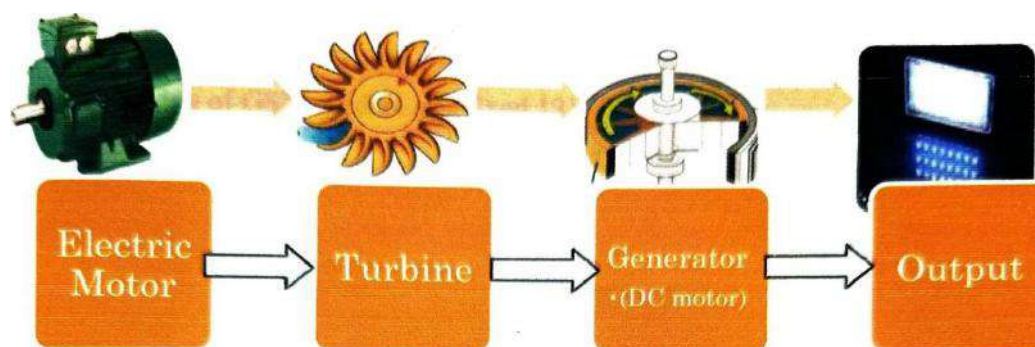
power generating projects are strongly advised, and the welfare of the agricultural community is the project's second most crucial component. Due to inadequate communication between farmers and on-field models, there is a high rate of power loss or damage to field machinery in remote locations. In this project, farmers and real-life models will be able to communicate wirelessly thanks to the GSM module.

#### LITERATURE SURVEY :

"Hydroelectric Power Project: An investigation of power schemes in Ravi basin" Dr. Mohan Kumar Slariya concentrated on the main study conducted in the state's Ravi basin. For a variety of causes, more than 80 power projects of varying sizes in 40 rivulets have been designated for planned development and have a negative influence on the body of traditional knowledge that is now available. Numerous experts have studied hydroelectric electricity while keeping in mind its benefits. Rising rates of economic expansion in the majority of emerging nations over the past few decades have led to increased levels of resource exploitation and environmental degradation, which, if left uncontrolled, might have a negative impact on future growth patterns.

Currently providing around 19% of the world's electricity needs, hydropower output is expected to triple over the next century. Nonetheless, research suggests that decreased river flows brought on by climate change will result in decreased hydropower generation. The need for electrical power has grown over the past year in the economy of the growing nation. This is true for both industrial and agricultural purposes. Assuming five KW per water mill, the 84 water mills in the region have a combined power output of 420 KW. In addition to grinding, it can generate 420 units of electricity per hour, which can be sold for Rs. 420 x per hour. This adds up to Rs. 24,192 per day, Rs. 72,576 per month, and Rs. 8,70,912 per year. Each water mill will directly employ at least three people ( $84 \times 3 = 252$  persons), and it will indirectly increase manifolds.

#### SYSTEM ARCHITECTURE :



*Figure 9 Block diagram for Power Generation*

The computerized architecture or operation of the corresponding project is represented by the accompanying diagram. The constant flow of water from the pipe causes the turbines to rotate when the motor is turned on. The blades of turbines tend to revolve because of the water's potential energy and the force with which it falls. This turbine produces energy when it is coupled to a generator that has a rotor and stator.

The generator's output of power has spikes and is not consistently stable. Therefore, an AC balancer is needed. Following the connection of the load, the stabilised power is stored in the battery. As waterfalls pass over the turbines, the force of the water causes the turbines to rotate, producing energy and voltage, which is shown on the LCD panel. There are several uses for the electricity that is produced.

The circuit diagram illustrates how several parts are connected, including an LCD, a GSM model, a PIC microcontroller, and crystals. ADC receives the generated voltage for the LCD

panel. The GSM module operates as a fixed 12V supply with the help of an external power supply.

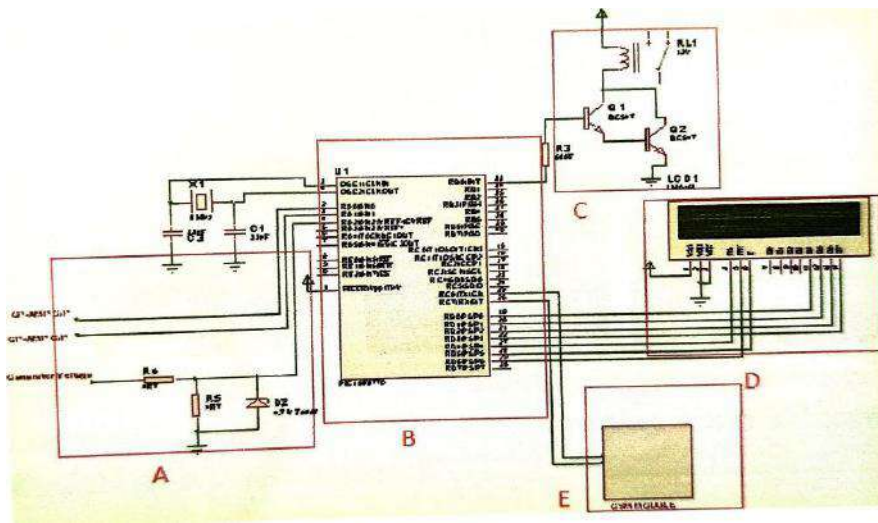


Figure 10 Circuit Diagram

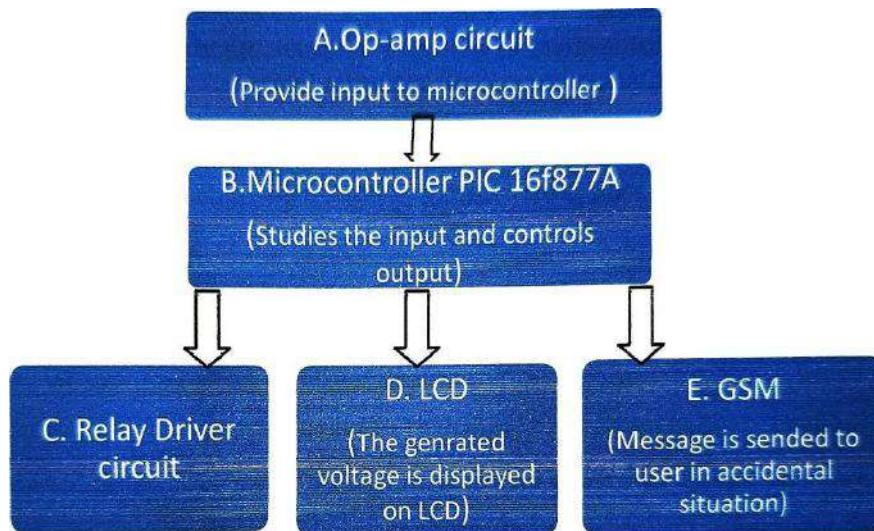


Figure 11 Flow Diagram

- A. **Input section:** As diagram shows the input to pic microcontroller is given from the output of op-amp circuit. Op-amp circuit is used for amplification purpose.
- B. **Microcontroller:** Microcontroller 328PU devices are available only in 28pin packages, while PIC16F874A/877A devices are available in 40-pin and 44-pin packages. The pin description of PIC 16F877A microcontroller is as follows:-
1. Pin no 2, 3 and 4 these pins are connected to op-amp to provide output.
  2. Pin 13, 14 for providing the frequency of oscillations.
  3. Pin no 19 to 29 these pins are connected as input to LCD from DO-DI.
  4. Pin no 25 and 26 connected to GSM module.
  5. Pin no 33 connected to voltage follower along with relay switch.
  6. Pin no 40 used as a ground
  7. Pin no 15 to 25 and 34 to 39 are unused.
- C. **Relay Driver Circuit:** A relay is an electro-magnetic switch which is useful if you want to use a low voltage circuit to switch on and off a light bulb (or anything else) connected to

the 220v mains supply. The current needed to operate the relay coil is more than can be supplied by most chips (op. amps etc), so a transistor is usually needed. A resistor of about 4k7 will probably be alright. Generally the diode is needed to short circuit the high voltage induced when current flowing through the coil is suddenly switched off.

D. **Liquid Crystal Display:** An LCD consists of two glass panels, with the liquid crystal material sandwiched in-between them. The inner surface of the glass plates are coated with transparent electrodes which define the character, symbols or patterns to be displayed. Polymeric layers are present in between the electrodes and the liquid crystal, which makes the liquid crystal molecules to maintain a defined orientation angle. When the LCD is in the off state, light rays are rotated by the two polarizers and the liquid crystal, such that the light rays come out of the LCD without any orientation, and hence the LCD appears transparent.

E. **GSM Module:** Output of PIC is connected to the GSM module and according to the behavior of motor observed by the controller the decision taken by the module and in any accidental case the SMS is given to the user

### **ADVANTAGES :**

- Small-scale hydropower is a clean energy source, producing no water or air pollution
- As a non-consumptive water use, small hydropower is a renewable energy source
- There is minimal impact on the environment
- As no reservoirs are created, small hydropower does not cause the problems associated with reservoirs such as methane emissions, displacement of people, sedimentation, and disrupted stream dynamics.

### **RESULT :**

In this experiment we have generated, stored and utilized the power, which is generated due to flow of water. The hydropower generated is 360 mW. Power used is been calculated as 180 mW, And the power stored is 180 mW.

### **CONCLUSION :**

Reclamation is helping to meet the needs of our country, and one of the most pressing needs is the growing demand for electric. This restructuring increases the importance of clean, reliable energy sources such as hydropower,

Hydropower is important from an operational standpoint as it needs no ramp-up time, as many combustion technologies do. Hydropower can increase or decrease the amount of power it is supplying to the system almost instantly to meet shifting demand. With this important load-following capability, peaking capacity and voltage stability attributes, hydropower plays a significant part in ensuring reliable electricity service and in meeting customer needs in a market driven industry. In addition, hydroelectric pumped storage facilities are the only significant way currently available to store electricity. Hydropower, besides being emissions-free and renewable has the above operating benefits that provide enhanced value to the electric system in the form of efficiency, security, and most important reliability. Water is one of our most valuable resources, and hydropower makes use of this renewable treasure. Small-scale hydropower is a clean energy source, producing no water or air pollution. Hence our project is efficient for this purpose. As most small hydropower schemes are integrated within an already existing water infrastructure, the effect on the local environment is minimal. Hence our project is non-polluting and renewable.

**REFERENCE :**

1. Theofanis P. Lambton, Christos C. Anastasia, Christos G, Panayiotou, and Marios M. Polycarpou, (2014), "A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking water Distribution Systems" *IEEE SENSORS JOURNAL, VOL 14, NO. 8, AUGUST*.
2. Pornjit Pratumsumwan and Watcharin Pongaen" *An Embedded PLC Development for Teaching in Mechatronics Education" IEEE- Mechatronics Educational Research Group, Teacher Training in Mechanical Engineering Department King Mongkut' s University of Technology North Bangkok (KMUTNB) Bangkok, Thailand.*
3. JYo-Ping Huang, Change-Tse Chou, Jung-ShianJau et.al (2010), "Water Quality Monitoring with Ubiquitous Computing" *IEEE*.
4. T. P. Lambrou, C. G. Panayiotou, and C. C. Anastasiou, (Oct. 2012)"A low-cost system for real time monitoring and assessment of potable water quality at consumer sites," in *Proc. IEEE Sensors*, , pp. 1-4.
5. New Delhi, India: Renewable Energy Cooperation- Net SNGH, D. (2009): *Micro Hydro Power Resource Assessment Handbook*
6. San Bruno, Choulot, Denis, V. (2010): *Energy Recovery in Existing Infrastructures with Small Hydropower Plants*
7. Guide on How to Develop a Small Hydropower Plant. Bensels, Belgium: European Small Hydropower Association (ESHA).
8. Slariya, Mohinder (2007), "A Study of Impacts of Hydroelectric Power Projects on Ecology and Society in Chamba District of Himachal Pradesh." *PhD, thesis submitted to MJP Rohilkhand University, Bareilly; Uttar Pradesh; India*



# **Computer Science & Application**



---

**1****A Study on the Fog- Edge-Cloud Computing based IoT (FECIoT):  
Architecture, Security, and Privacy Issues****Prof .Ather Iqbal**Department of Computer Science  
Vidyabharati Mahavidyalaya Amravati  
atheriqbal13@gmail.com**Dr.C.H.Sawarkar**Department of Comp.Science  
Narsamma Mahavidyalaya,Amravati  
chsawarkar@gmail.com**Dr.Shilpa S. Sarvaiya**Department of Comp.Science  
VBMV,Amravati  
sarvaiya.shilpa@gmail.com**ABSTRACT**

The Internet -of -Things (IoT) is the future of the Internet, where everything will be connected. Studies have revealed that Fog-Edge-Cloud Computing (FEC) –based services will play a major role in extending the cloud by carrying out intermediary services at the edge of the network. Fog-Edge-Cloud Computing-based IoT (FECIoT) distributed architecture enhances service provisioning along the Cloud-to-Things continuum, thereby making it suitable architecture. Furthermore, the proximity of Fog-Edge devices to where the data is produced makes it stand out in terms of security and privacy issues. From the business perspective, FECIoT will lead to a boom and spring up of Small-to-Medium-Sized enterprises (SMEs), thereby encouraging inclusion for all. In this paper present a study on FECIoT.

**Keywords**— Fog-Edge-Cloud Computing (FEC), Internet-of-Things (IoT), Service Oriented Architecture (SoA), Cloud-to-Things (CoT), Attacks.

**I. Introduction**

Recent studies have shown the shortcomings of the cloud as regards handlings of big-data. By the year 2020,it is projected that about 50 billion things are expected to be connected to the Internet [1].To this point,IoT requires a robust and resilient architecture that will enable faster data processing, as well as storage. Several researchers have suggested the need to integrate the Fog-Edge Computing (FEC) with the IoT [2].FEC promises to run IoT-enabled applications for real-time control and analytics, with millisecond response time. Furthermore, FEC enables designing and building a scalable and adaptable IoT platform. A service-based architecture (SoA) is a component-based model that focuses on the systematic design of the workflow of coordinated services. Where and how to perform computation and storage along the Cloud-to-Things (CoT) continuum, and how decisions can be managed within heterogeneous systems is still a debatable issue [3].Localized data analytics coupled with control can provide some level of autonomy to devices close to the edge of the network (FEC devices), which may help in enhancing the performance of mission-critical IoT applications. This paper presents new challenges in emerging IoT and the bottlenecks faced in resolving these challenges using today's computing and networking models. The paper further discusses why the FECIoT architecture should be deployed to fill possible technological gaps with a view to enhancing new business opportunities. Furthermore, we discuss security features, as well as security challenges that exist within the FECIoT framework. The remainder of this work is organized as follows. In section II, we present the basic concepts. The FECIoT architecture framework is presented in section III.Security and privacy issues are presented in section V.

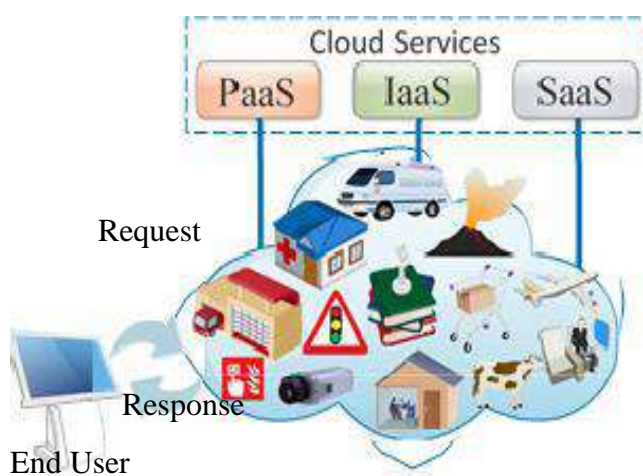
**II. Overview of Basic Concepts**

In this section, present basic concepts and provide comparisons amongst seemingly similar concepts.Thus, we provide an overview of Cloud computing, Fog Computing, Edge computing, similarities and differences [4].



### A.Cloud Computing (CC):

With top multinational computing giants like Amazon web services, Microsoft Azure, Google Cloud platform, and IBM Cloud championing the adoption of the generic cloud computing model where big-data analytics, decision making, and computations all take place centrally in the distant cloud data-centres. The increase in Machine Type Communication (MTC) as observed in IoT will lead to massive amounts of data flow within the IoT ecosystem. As such it becomes difficult to manage traffic and congestion within the network using the CC model. Despite the emergence of FEC which promises better business prospects for SMEs, lower latency, and higher bandwidth efficiency, the cloud will continue to have a key role to play in the proposed FECIoT framework. Figure 3 shows the Cloud Computing model along with three generic services. These services are Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS), and Software-as-a-Service (SaaS).



**Figure1: Cloud Computing Model.**

**I) Platform-as-a-Service (PaaS):** This is a customer-based Cloud Computing service that supports clients, by given them the flexibility of developing, running, and managing Web based without going through the rigour of building and maintaining infrastructures typically associated with developing and launching an application. The PaaS also supports the overall life-cycle management of cloud applications, including coding, testing, deployment and maintenance [5]. A good example is Apprenda, which is a provider of private cloud PaaS for .NET and Java.

**II) Infrastructure-as-a-Service (IaaS):** This service model is also known as Hardware-as-a-Service (HaaS), it is a cloud computing service model that provides computing infrastructure to enhance enterprise operations, usually based on outsourcing. In other words, IaaS manages computing, storage and networking resources and provides basic resource services to the PaaS or to users directly [7]. Generally, IaaS provides hardware (may include software), storage, servers, and data center space or network elements. Amazon Web Services (AWS), Cisco Metacloud (formerly Metapod), Microsoft Azure, Google Compute Engine (GCE), and Joyent etc. all fall within the IaaS category.

**III) Software-as-a-Service (SaaS):** This is a software distribution model which allows clients to have access to applications hosted by third-party providers over the Internet [6]. Good examples of SaaS used in everyday life are Twitter, Instagram, Facebook, and Google's suite of intelligent apps (formerly Google Apps).

---

## **B. Fog Computing (FC):**

The concept of fog computing (FC) was first introduced in 2012 [7], working at Cisco Inc. The FC paradigm entails moving intelligence down to Local Area Network (LAN) level and data is processed at an IoT gateway. The main aim of its introduction was to extend services and functionalities offered by the cloud at the edge of the network. Such functionalities may include storage, processing, database operation, integration, security, and management to IoT end-devices, leveraging on its proximity to the edge of the network. With exciting benefits of minimizing network congestion, minimizing end-to-end latency, tackling connectivity bottlenecks, improving security and privacy, and enhancing scalability, FC is seen as the way forward. Furthermore, there are claims within the industry of the vast business opportunities that could be derived with the advent of FC. With the effective distribution of computing, storage, networking, and management service along the Cloud-to-Things continuum, it meets today's application requirements for local content, resource pooling, and real-time processing [8]. As such, FC has attracted interest from both the academia and industry. It is a fact that FC does not replace the CC, rather it complements by offloading data or service request that can be processed locally. However, we acknowledge the limitations of the CC-based model and stress the need for FC integration to allow for global applicability. Another emerging paradigm that can revolutionize IoT is Dew Computing (DC). In, DC is expected to depend on micro-service approach in a highly heterogeneous, vertical, and distributed hierarchy. It gives room for a centralized-virtualization-free computational horizon where data scattering into low-end devices is possible. Hence, allowing for data accessibility even without continuous internet access. The extreme scalability and self-adaptive attributes of DC makes it prime to the success of IoT [9]. With FC's intermediary role of deploying existing computing infrastructure in bridging the cloud to things, FC will be prime to the success of existing and emerging technologies like the smart grids, smart homes, smart cities, wireless sensor networks, mobile healthcare, manufacturing, vehicular networks, and lot more. Below are some advantageous features of FC:

- 1) Geographically dispersed.
- 2) Support for large-scale sensor networks and end nodes.
- 3) Provides better real-time response than the Cloud-based model.
- 4) Online analytics and interaction with the cloud.

## **C. Edge Computing (EC):**

As the name implies, the edge computing (EC) entails computation that is carried out at the edge of the network EC aims at overcoming limitations associated with the cloud computing-based model. It serves as the intermediary between the end users/devices and the cloud, providing processing and storing functionalities to a large number of IoT end-devices. The proximity of edge devices will minimize computational load on data centers situated far away in the cloud. Real-time response will be enhanced, as well as reduced latency [10]. Another merit of EC is the distributed nature and support for device mobility within heterogeneous networks. According to edge layer can be implemented in three modes, the MEC, FC, and Cloudlet Computing (CC). Hence enabling cloud computing functionalities inside the Radio Area Network (RAN). Cloudlets is a smaller version of the cloud which uses dedicated devices that offer cloud like functionalities. Below are some advantages features of EC:

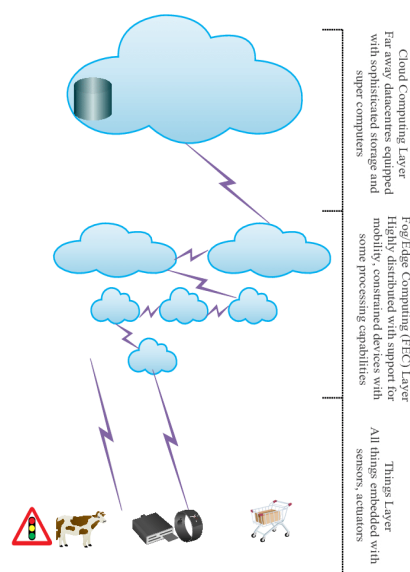
- 1) Geographically dispersed.
- 2) Improved security, as encrypted data moves further into the core network.
- 3) Provides better real-time response than the cloud based model.
- 4) Better scalability through virtualization.
- 5) Limits potential communication bottlenecks.

### D.Fog-Edge Computing (FEC):

It is pertinent to note that FC devices may not necessarily be at the network edge, but reside close to the edge of the network. In contrast, edge devices often reside at the network edge, and are often the first point of contact to the IoT end- devices. In essence, FC devices and EC devices are both close to the IoT end-devices, but the EC devices are often closer. In many works, fog computing and edge computing have been used interchangeably. Some consider FC as a part of the EC and micro data center (MDC) paradigm for IoT. Both FC and EC have their services located close to the end users, however, EC is resident in edge devices, while FC resides in the network edge devices, usually a single or few network hops away from the edge. The EC platform has constrained energy and limited storage, and fall within the class of constrained devices. The increase in the number of IoT applications may result in higher contention for resources and additional latency [11]. In essence, the resource contention of EC is greater than FC due to proximity to IoT end-devices. Furthermore, EC focuses more on the things domain, while fog computing focuses more on the infrastructure domain. FEC has certain pillars, they include security, scalability, openness, autonomy, reliability, agility, hierarchical organization and programmability, which is inherent to both FC and EC. As such, the motivation for integrating both the FC and EC is based on the peculiarities between them.

- 1).They both use a virtualized IaaS platform and allow for multi-tenancy of applications at the edge of the network.
- 2) They both compliment functionalities offered by the cloud and are located between end users and data centers.
- 3) They both can be physically co-located with access points, roadside units, base stations, routers, switches, and gateways.
- 4) They both are mostly deployed wireless and provide low latency, low jitter, and cognition within the system.
- 5) They both provide computational services in distributed geographical locations in order to minimize the load on the cloud.

Figure 2. Shows a pictorial representation of the FECIoT model with various domains. In this paper, we arrive at a consensus that FC and EC are congruent [12].



**Figure 2: FEC Architecture and Interaction in the Cloud-To-Things Continuum**

### III. FECIoT Architecture:

The term "FECIoT" was first coined by Lin et al. in [13] with a motive to emphasize the immense potential that could be derived when the Fog-Edge computing paradigm is well integrated into the IoT architecture. In this section, present the FECIoT architecture framework. Fog-Edge devices may be linked to form a mesh to provide load balancing, resilience, fault tolerance, data sharing, and reduction in the Cloud-to-Things communication. Architecturally, this demands that Fog-Edge devices have the ability to communicate both vertically and horizontally within the IoT ecosystem. The FECIoT inherits the basic IoT architecture and delivers all IoT requirements in a more efficient way by leveraging on the distributed FEC paradigm. In this paper describe three different architectures that are Three-layer, Four-Layer and Five-Layer Architecture.

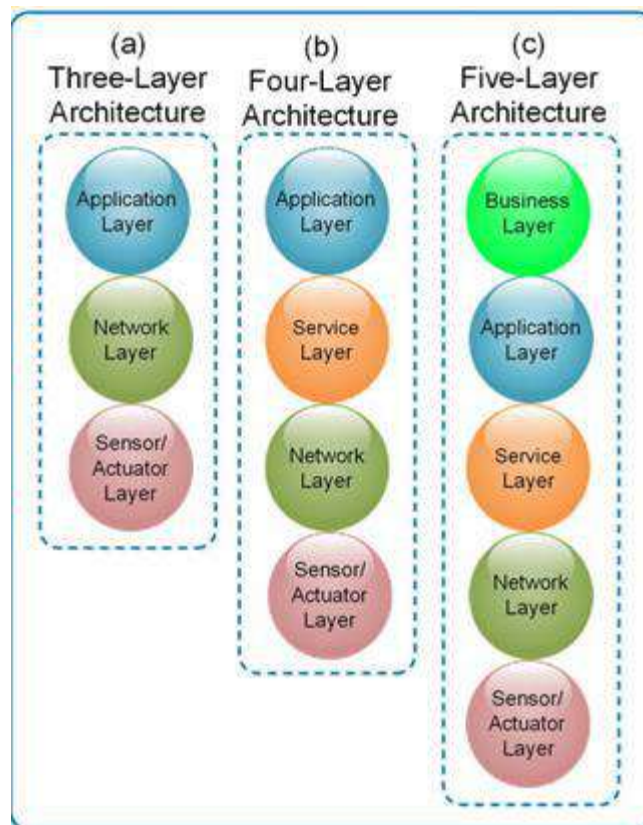


Figure 3: FECIoT (a) Three-Layer Architecture, (b) Four-Layer Architecture, (c) Five-Layer Architecture

#### A. Three-Layer Architecture:

This architecture is traditional in design and considers three basic layers. Figure 4 (a) shows the Three-Layer Architecture comprising of: (I) The Sensing Layer, (II) The Network Layer, and (III) The Application Layer.

**(I) Sensing Layer:** This layer is also known as the Perception Layer, acquires data through sensing using Radio Frequency Identification (RFID), Sensors. Nodes within radio range of each other may collaborate for the purpose ensuring ubiquitous data communication within the IoT network.

**(II) Network Layer:** This layer performs the task of routing data across diverse networks. Sensed information is received from the sensing layer and then routed to IoT hubs and devices over the Internet. This layer supports computing platforms such as Cloud Computing platforms,

Internet gateways. Which operate using state-of-the-art technologies like 5G/LTE, Bluetooth, WiFi. The Network Layer uses gateways to send data to and from applications or things across heterogeneous networks, and over multiple protocols and technologies.

**(III) Application Layer:** This layer provides services (Storage, Processing or Analysis) based on received data or request from the Network Layer. Several IoT applications exist in this layer with diverse requirements and deployed together with the Middleware functionalities. With emerging Fog-Edge deployments, Multivendor ecosystem applications need to be able to migrate and operate seamlessly despite system heterogeneity.

The Three-Layer Architecture looks simple, however, when taking a closer look at the Network and Application Layer, we observe complexities in grafting data services (Data Aggregation, Data Mining and Analytics) into this architecture. Thus, giving rise to a new layer called the service Layer.

## **B. Four-Layer Architecture:**

This Architecture is also known as the Service-Oriented Architecture (SOA). The SOA is the application framework that allows establishments and enterprises to build, deploy and integrate these services independent of the technology systems on which they run. Here, the service layer is placed in between the Application and Network Layer in order to enhance data Services in IoT.

This Service-Oriented Architecture focuses on designing the work-flow of coordinated services, and allowing for hardware/software reuse because it supports the design, deployment, and integration of services, which are not dependent on the technology platform they operate on [18]. Figure 3 (b) shows the Four-Layer Architecture comprising of: (I) Sensing Layer, (II) Network Layer, (III) Service Layer, and (IV) Application Layer, which are briefly discussed with emphasis on the Service Layer.

**(I) Sensing Layer:** This Layer is at the base of this architecture, and responsible for data collection, measurement and extraction of physical devices. This data is passed on to the upper layers.

**(II) Network Layer:** This Layer provides support for data to be transmitted over multiple networks and topologies. Route decisions are made in this layer.

**(III) Service Layer:** This Layer as the name implies, provides a variety of services. This layer is also known as the Interface or Middleware Layer. The Service Layer can be further broken down into four components, namely:

- 1. Service Discovery:** This helps in discovering desired service request. In a global Service Discovery framework was introduced which allows users to register their own sensors into a common infrastructure, and discover the available services via a mobile device.

- 2. Service Composition:** A sub-layer in the SOA which provides functionalities for the composition of specific services offered by networked objects in building specific applications. Web of services also plays a vital role, in the sense that they allow for a precise definition of capabilities of interfaced objects and interaction with them.

- 3. Service Management:** This provides the primary functional requirements and management for each object. The functionalities in services may cut across QoS management, lock management, and semantics. In addition, newer services may be deployed at run-time, in order to meet application requirements.

- 4. Service Interface:** This interface serves as a bridge to connect all provided services. Interfaces are necessary for the reduction of complexities in business processes

**(IV) Application Layer:** This Layer is at the top of this architecture, providing overall support based on system's functionalities to end user. Unlike the traditional Three-Layer Architecture, the Application Layer is not part of the Middleware, rather it instructs the Service/Middleware Layer. This Layer provides an interactive interface via standard Web Service Protocols and

Service Composition Technology over heterogeneously distributed systems and applications. Examples of such applications include, smart homes, intelligent transportation, smart industry, smart health-care.

**C. Five-Layer Architecture:** This model has a Business outlook and is extracted from the traditional Application Layer to provide more complex services. The Five-Layer Architecture comprises of the following: (1) Sensing Layer,(2)Network Layer,(III) Service Layer,(IV) Application Layer, and ( V) Business Layer.Here,we focus on the Business Layer, since previous Architecture has covered lower Layers.

**(V) Business Layer:** The main role of this Layer is to record and analyse all IoT CPS (Cyber Physical Systems) operations. The business Layer handles the entire IoT System, which includes applications, business and profit models and user confidential information. Figure 4 (c) shows the business Layer as an additional feature in the SoA.The SoA actually facilitates the creation of systems, which support the derivation of independent business solutions from technological constraints .The FEC Architecture will play an important role in reshaping the networking, server and software industry, with the convergence of routers,switches,storage and application servers into FEC devices.Furthermore,the distributed FEC Architecture supports the emerging Fog-as-a-Service ( FaaS),where smaller business enterprise can also participate in delivering Private and Public services at diverse scales to end user.

## IV. Security and Privacy in FECIoT

Security is important for the safe and reliable operation of connected IoT devices. With pervasive data emanating from heterogeneous systems, data confidentiality, integrity, privacy, as well as authentication to verify data source is very important. However, due to processing limitations of IoT devices, it is almost impossible to deploy full-fledged security suites. The FECIoT helps in overcoming some of the challenges encountered in existing IoT architectures that use the cloud computing-based model. In the aspect of security, FEC devices can be deployed as proxies for IoT end-devices. Despite the merits of FECIoT, there exist several security issues associated with this architecture. In this section, we present some security features of FECIoT and possible security attacks in FECIoT [13].

### A.Security Features in FECIoT:

Security requirements impact different layers depending on the specific security principles. We briefly discuss some important security features in FECIoT.

**1) Trust:** Establishing communications between IoT devices, FEC devices, and infrastructures in the cloud requires some level of trust. Also, to effectively implement trust within the FECIoT architecture, devices need to be equipped with adequate security, making them trusted elements. When trusted devices are deployed, it provides the basis for a secure FECIoT ecosystem. Authentication and transparency play an important role in fostering prior relations between devices. Trust is not limited to communications among devices, rather covers the relationship between different IoT layers and applications. Several trust-management models have been applied in cloud computing domain, using artificial intelligence, fuzzy methods, game theory, and Bayesian estimation- based techniques.

**2) Authentication:** This involves entity identification. Before a device can become part of any given network, it is necessary that the device is first authenticated. However, the constrained nature of IoT devices makes it even more challenging when considering complexity in both registration and re-authentication phases. FEC-based authentication servers will be a better choice for the centralized cloud authentication servers, due to the distributed nature and proximity of FEC devices [14].

**3) Integrity:** Cannot be altered during the process of data transmission. Integrity is assured only when the intended and authorized entity receives data accurately as was sent.

Compromised data may cause serious disruption within the network and further cause harm to the operation of the IoT application. In, a sampling and signature scheme was presented, providing opportunity to relieve the burden of the network, where the local collector acts as the coordinator and periodically transmits the sampled packets to the global traffic analytic. This scheme was able to provide integrity and can be modified to suit the FECIoT framework. A Game-theoretic approach was adopted in to examine the best strategies to slowly corrupt the integrity of an IoT network. This approach can be used in designing better defensive measures in FECIoT.

**4) Confidentiality:** Confidentiality ensures that only authorized users/devices can have access to useful information or modify it, hereby keeping unauthorized users/devices away from interfering with data and services. Data in the FECIoT framework flows from the physical devices (e.g. sensors and actuators) through to FEC devices and then to/from higher layers. This increases the chance for this data to be accessed by malicious devices within the network. It is pertinent to address the access control mechanism and also the device authentication process.

**5) Privacy:** Privacy ensures that data is accessed only by the corresponding entity/device within the network. It is important to ensure that other users/devices may possess some specific controls based on received data, but should be unable to infer other useful information from the received data. Due to the huge number of IoT end-devices, and sheer volume of data flowing within the FECIoT ecosystem, privacy cannot be undermined

**6) Availability:** Availability is a very crucial security feature in FECIoT. It ensures that data and system resources should be available to authorized users/devices requesting for data or services. Most IoT applications are latency-sensitive, and as such, any downtime in system operations may have an adverse impact on end-users. Distributed denial of service (DDoS) attack is one that renders data and services unavailable to legitimate users/devices.

**7) Access Control:** Access Control is the process of determining whether user/devices can have access to system resources, this could be data, or services. This process involves denying or revoking access, especially to unauthorized users/devices. In, an access control system was developed which enables offloading of complex access control decisions to third, trusted parties. The design which is based on a simple communication protocol imposes minimal overhead. Thus, making it suitable for FECIoT applications.

## **B. Possible Security Attacks in FECIoT:**

Here, we present possible security attacks in FECIoT [14].

**1) Distributed Denial of Service (DDoS):** One of the most lethal attacks in the FECIoT architecture is DDoS. The risk of malicious clients and coordinated group of clients (Botnets) mounting DoS attacks is still an issue of concern. DDoS attacks may emanate from IoT end devices. On the other hand, FEC devices may also be used to launch DDoS attack.

**2) Man-in-the-Middle Attack (MitMA):** The Man-in-the-Middle Attack is a prominent attack that could constitute a serious threat in FECIoT, especially in the area of privacy. The attack easily exploits this platform to disclose sensitive information such as location and identity of the FEC devices. This kind of attack is often successful, as devices cannot implement secure communication protocols due to resource constraints this attack still poses a serious challenge in FECToT.

**3) Physical Attack:** This type of attack involves physical compromise of hardware components. This hardware components could be RFID tags, sensor devices, FEC devices, or even more centralized infrastructure. Susceptibility of this kind of attack varies with respect to the location of deployment, level protection given to such devices.

## V. Conclusion

FECIoT has the potential to add value to existing IoT systems by enabling real time response as well as providing storage and computational services in a distributed manner to IoT end devices. The proximity of FEC devices to where the data is produced makes it stand-out in terms of resource allocation, service delivery, and privacy. The FECIoT framework offers more responsiveness and eliminates the need for costly bandwidth additions by offloading gigabytes of network traffic from the core network. The proposed FECIoT service-based framework will greatly enhance service delivery to IoT end-users, hence, FECIoT should be considered as part of the overall Internet of the future, which will transform the Internet industry. In this paper, study the key aspect of the FECIoT framework and presented security and privacy issues. Also analyse FECIoT in a comprehensive form, especially for new entrants in the area of IoT. It should, however, be noted that the FECIoT framework may seem to have provided improvements on existing frameworks, numerous security and privacy issues abound. The FECIoT promises better service delivery to end users, and inspire novel business models. This is expected to be a prime focus for researchers in the next decade.

## VI. References

- [1] Ms.S.B.Sarvaiya, Dr.S.E.Tayde, "Analysis of Security Mechanisms Based on IoT, Fog and Cloud Computing Paradigm." 7TH National Conference on Recent Trends in Computer Science & Applications (RTCSA-2018) 18 & 19 Dec 2018, ISSN: 2249-894X.
- [2] Dr.S.E.Tayde, Ms. S.B.Sarvaiya,"The Merger of Cloud Deployment Models Services With IoT.", NCETS "Research Journey" International E- Research journal, Impact Factor 6.261 ISSN: 2348-7143,February-2019.
- [3] Dr.S.E.Tayde, Ms.S. B.Sarvaiya," The Role of Just-in-Time Indexing Technique of IoT on Cloudlet-based during Interactive Data Exploration System", Recent Advances in Science and Technology (RAISAT-2019), 5 and 6 March-2019.
- [4] Dr.S.E.Tayde, Ms. S.B.Sarvaiya," EDGE CLOUD COMPUTING COMPLIMENTARY ROLE IN IOT ENVIRONMENT", 3<sup>rd</sup> National Conference Recent Development Science, Engineering and Technology (RDSET 2019), April 03, 2019.Vol.4, Issue 6, ISSN-2349-5162.
- [5]A.Chen, H.Wu, L.Tian, G.Luo," HCOS: A Unified Model and Architecture for Cloud Operating System", ZTE Communications Magazine, Vol.15, No.4, PP.23-29, NoV.2017.
- [6] W.Shi, J.Cao, Q.Zhang, Y.Li, L.Xu,"Edge Computing: Vision and Challenges", IEEE Internet of Things Journal, Vol.3, No.5, PP.637-646, OCT.2016.
- [7]J.Pan, J.McElhannon,"Future Edge Cloud and Edge Computing for Internet of Things Applications." IEEE Internet of Things Journal, Vol.5, No.1, PP.439-449, Feb.2018.
- [8]S.Chen, T.Zhang, W.Shi, "Fog Computing", IEEE Internet Computing, Vol.21, No.2, PP.4-6, Mar.-Apr.2017.
- [9]P.P.Ray,"An Introduction to Dew Computing: Definition, Concept and Implications", IEEE, Access, Vol.66, PP.723-737, 2018.
- [10]K.Dolui, S.K.Data," Comparison of Edge Computing Implementations: Fog Computing, Cloudlet and Mobile Edge Computing", IEEE Global Internet of Things Summit (GIoTS), Geneva, PP.1-6, 2017.
- [11] M.Aazam, E, N.Huh,"Fog Computing: The Cloud-IoT/IoE Middleware Paradigm", IEEE Potentials, Vol.35, No.3, PP.40-44, May 2016.
- [12]P.Hu, S.Dhelim, H.Ning, T.Qiu,"Survey on Fog Computing "Architecture, Key Technologies, Applications and Open Issues", Journal of Network and Computer Applications, PP.27-42, Nov.2017.
- [13]J.Lin,W.yu,N.Zhang,X.yang,H.Zhang,W.Zhao,"A Survey on Internet of Things ( IoT ) Architecture, Enabling Technologies, Security, Privacy and Applications",IEEE Internet of Things,Vol.4,No.5,PP.1125-1142,Oct,2017.
- [14]H.Kim, E.A.Lee,"Authentication and Authorization for the Internet of Things (IoT)", IT Professional, Vol.19, No.5, PP.27-33, 2017.



## IoT Node Security Attacks on Device Layer: Attacks Detection Countermeasures and Solutions

**Dr. Shilpa B. Sarvaiya<sup>1</sup>, Dr. D. N. Satange<sup>2</sup>, Dr. A. A. Tayade<sup>3</sup>**

<sup>1</sup>Department of Computer Science, Vidyabharati Mahavidyalaya, Amravati, sarvaiya.shilpa@gmail.com

<sup>2</sup>Department of Computer Science, Narsamma Hirayya Arts Com & Sci, Amravati  
dineshnsatange@rediffmail.com

<sup>3</sup>G. S. Science, Arts and Commerce College, Khamgaon, email : arvindtayade40@gmail.com

### ABSTRACT

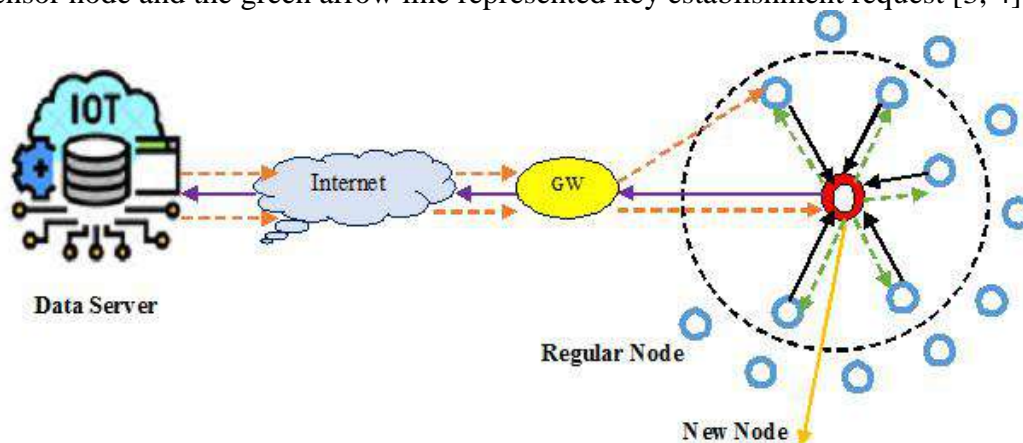
Node identity authentication is an essential means to ensure the security of the Internet of Things. If there is any attacks perform in IoT network then due to the authentication the intruders can be trap and system will not allow to change the network data. Trusted IoT node providing existing or new node authentication and authorization for network credentials, extract node data for extracting IP address, form packet to secret key generation for data privacy. Thus, secure authentication is a major requirement for managing and communicating with respect to the devices in the IoT environment. Currently most IoT devices use default login credentials and not secured with better configurations and protocols which paves way for various types of attacks. In this paper we address the physical node capture attacks, attacks detection on device layer or physical layer, corresponding countermeasure and finally suggesting their solutions for stability under these attacks in IoT network.

**Keywords:** Attacks, Authentication, Authorization, Countermeasures, Device layer, Internet of Things, IoT Network, IP address, packet generation, Trusted Node, secret key, Security.

### 1. Introduction

Figure 1 demonstrates IoT node enabled architecture is an important part of the IoT network. In order to protect the security of IoT extract node data approach provides a convenient way to secure end to end communication environments and allows numbers of nodes to establish a secure channel with the help of the trusted server [1, 2].

In figure 1 the dash line circle denoted the transmission range, the blue concentric circle denoted the regular node, and the red concentric circle denoted the new node. The black arrow line represented the message from one node to their neighbour node; the purple line represented the integration message for the data server, red dash line represented the secure key message for sensor node and the green arrow line represented key establishment request [3, 4].



**Figure 1: IoT Node Enabled Architecture**

## 2. IoT Node Authentication Process

A secure communication mechanism is established by IoT nodes and devices; while their connection should be activated through the process of node and device binding in figure 2.

Our implementation given in figure 2 has following step by step approach to solving a task:

**Steps 1:** To create a node and add number of devices for every node with IP address to border router [5, 6].

**Step 2:** Perform device as well as node login in order that single authenticated user can see the password and connect with the device this procedure is called as IP based manual identification. [7, 8].

**Step 3:** Every node is configured with unique username and strong password.

**Step 4:** The particular node is going to be allotted with Node ID to particular device so, we can view the device [9, 10].

**Step 5:** The data will be considered for authentication of credentials.

**Step 6:** After that in authentication phase also checked the external or internal attacks present on device layer or physical layer of IoT network [11, 12].

**Step 7:** Further more stored the data in the cloud generated by the IoT device.

**Step 8:** Extract the IP address of IoT device form packets in a given node bind with the signal and send to receiver.

**Step 9:** The dashboard represent the list of nodes and its connected devices.

**Step 10:** The dashboard also help to control all nodes transmission and their data.

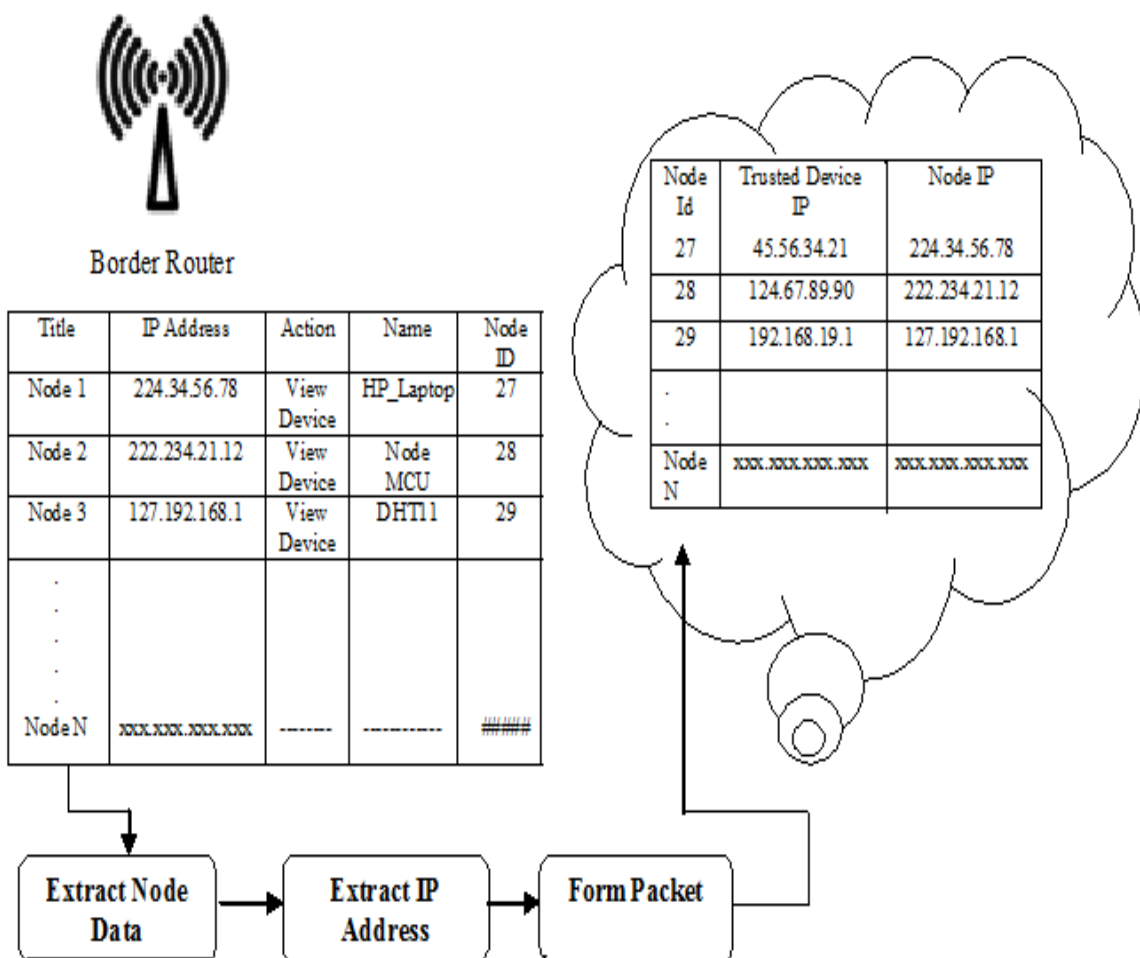


Figure 2: Secure IoT Node Schema

### 3. Attacks Detection and Problems formulation

Registered the IoT node for particular IoT device with unique node identification number in the process of heavy data streams are generally under security threats because of the following reasons [13, 14]:

- Unlatched Vulnerabilities.
- Lack of adequate Security Solution.
- Unchanged or Unsafe Passwords.
- Limited Memory.
- Limited Radio Bandwidth.
- Highly inefficient unsecured data transfer over the IoT devices network.
- Irregulars' updates and recovery.
- Undefeated Architecture.
- Insufficient memory with limited bandwidth.
- Poor network quality due to network congestion.

#### 3.1 Basic Attack Model:

Figure 2 describes that the intention of an attackers is to find some IoT vulnerabilities from the underlying IoT network and takes benefits of it to steal sensitive data from an IoT device.

IoT devices and IoT nodes are gaining more attention in the perspective of security implementation since it became of urgent need due to unauthorized access to its sensitive data for personal benefits [15, 16].

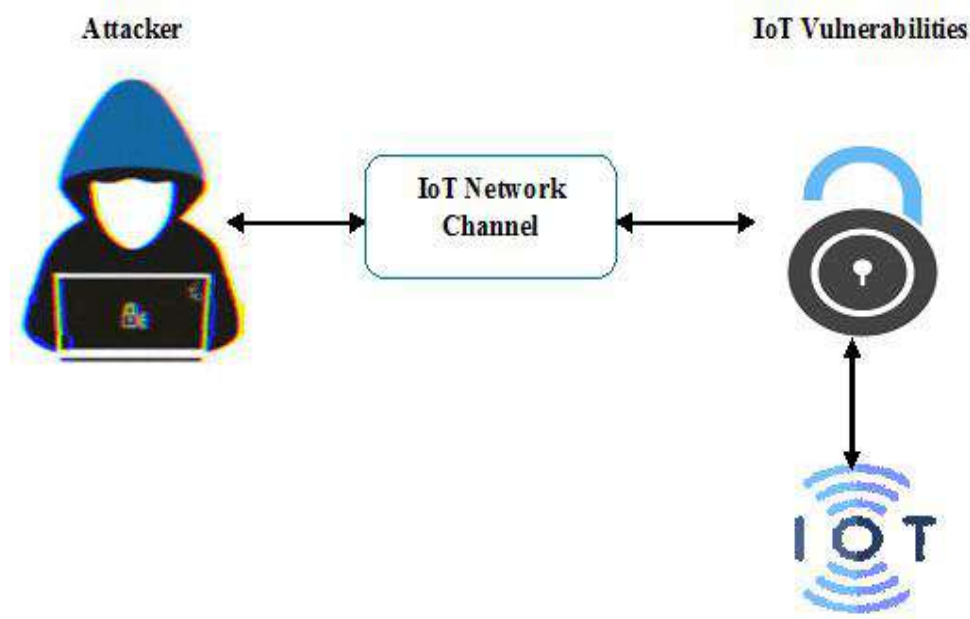


Figure 3: Basic Attack Model on IoT

#### 3.2 Device and Node Capture Attack Model

Here discuss the problem of physical devices and nodes capturing attacks model on IoT network. Extract particular node data with corresponding device name to form a sender-receiver verification and encryption-decryption scheme. The data in IoT network is remain secured and will also check for the sender and receiver terminal. If the sender is right and receiver is also right the cross verification of sender and receiver done as well [17, 18].

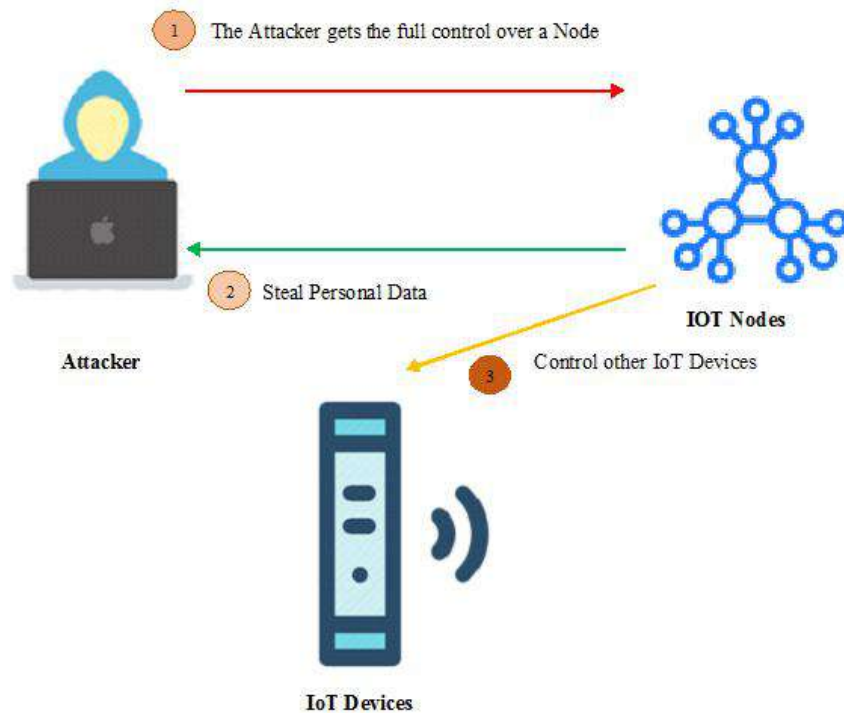


Figure 4: Devices and Nodes Capture Attack Model on IoT

#### 4. Attacks Countermeasures and Solutions

Security is defined as a process to protect a resource against physical damage, unauthorized access by maintaining a high confidentiality and integrity of the useful information and making information about that object available whenever needed. IoT is relying on connectivity of infinite devices for its operation [19, 20].

Hence, the possibility of being exposed to a security attack is most probable. Here classified the IoT security attacks and the proposed countermeasures based on the current security threats, considering the Device layer or physical layer or perception layer. Table 1 summarizes the taxonomy of attacks and feasible solutions of IoT nodes and devices under device layer [21, 22].

Table 1: Taxonomy of attacks and solutions in IoT device layer

Layer/Component	Attacks	Countermeasure	Solutions
<b>RFID Nodes</b>	DoS, Eavesdropping, Spoofing.	Secure localization, Privacy protection.	Access Control , Cryptography Techniques.
<b>Sensor Nodes</b>	Node Subdivision, Node Failure, False Node.	Passive Information Gathering, Don't Form traffic Collision.	Device and Node Authentication Process.
<b>Sensor Gateways</b>	Protocol Tunneling, Man-In-The-Middle, Signal Lost.	Maintain Device and Node IoT Board Security.	Integration Security Message security, Sensor Privacy.

## 5. Conclusion

This research finding on security risks in IoT emphasize the extension of the attack surface of the IoT threats and Vulnerabilities rise robustly as the connected IoT devices and nodes. Here IoT nodes create dynamic topology and perform their tasks without human intervention against various attacks. IoT devices in all aspects of human life indicate the necessity of considering these security threats before the implementation of the countermeasures. This paper also highlights the proper attacks detection and their suitable solutions, which have proved to be effective in securing communication between IoT nodes and devices. The authenticated binding process can be able to tackle the various attacks in IoT nodes and devices make the IoT network to strong so that this mechanism will help in future to integrate with the various types of projects and the leakage found in the services can be identify and rectify as per the concerning security algorithms and protocols. This studied have limitation in only node failure cases but the occurrence of this is to very less. Therefore, the future work will focus on improvement in dynamic performance of IoT node environment and the behaviour prediction of attackers.

## 6. References

- [1] Shiqiang Zhang, et al., "A Heterogeneous IoT Node Authentication Scheme Based on Hybrid Block chain and Trust Value", KSII Transactions on Internet and Information Systems Vol.14, No.9, PP.3615-3638, ISSN:1976-7277, September 2020.
- [2] Mohamad Faiz Razali, et al., "TPAL: A Protocol for Node Authentication in IoT", Journal of Computer Science, Vol.14, No.10, PP.1401-1411, doi:10.3844/jcssp.2018.1401-1411, 26 October 2018.
- [3] Chun-Ta Li et al., "A Secure Three Party Node Authentication and Key Establishment Scheme for the Internet of Things Environment", Journal of Internet Technology, Vol.19, No.1, PP.147-155, doi:10.3966/160792642018011901014
- [4] Ilker Yavuz, et al., "End to End Secure IoT Node Provisioning", Journal of Communications, Vol.16, No.8, PP.341-346, doi:10.12720/jcm.16.8.341-346, 8 August 2021.
- [5] V.S. Saranya, et al., "An Intelligent IoT Attack Detection Framework Using Effective Edge AI Based Computing", Indian Journal of Computer Science and Engineering (IJCSE), Vol.13, No.4 PP.1156-1167 E-ISSN: 0976-5166, P-ISSN: 2231-3850, July-August 2022, doi:10.21817/indjcse/2022/v13i4/221304059.
- [6] K. Ravikumar, et al., "Detection of Node Capture Attack in Wireless Sensor Networks", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue 4, PP.56-61, E-ISSN:2320-7639, 31 August 2018.
- [7] C. Ramakrishna et al., "A Survey on Various IoT Attacks and its Countermeasures", International Journal of Engineering Research in Computer Science and Engineering, Vol.5, Issue 4, PP.143-150, ISSN: 2394-2320, April 2018.
- [8] Bichen Che et al., "KNEMAG: Key Node Estimation Mechanism Based on Attack Graph for IoT Security", Journal of Internet of Things, Vol.2, No.4, PP.145-162, ISSN: doi: 10.32604/jiot.2020.010035, 15 August 2020.
- [9] Nivedita Sharma, et al., "Novel Technique for Detection of Malicious Nodes in IoT", International Journal of Research In Electronics and Computer Engineering (IJRECE), Vol.7, Issue 3, PP.483-487, ISSN: 2393-9028(Print), ISSN: 2348-2281(Online), July-Sept 2019.
- [10] Moinul Hossain, et al., "Hidden Terminal Emulation: An Attack in Dense IoT Networks in the Shared Spectrum Operation", Vol.1, ISSN:6, ISSN:7281-0962, 2019 IEEE.
- [11] Xin Liu, et al., "Identifying Malicious Nodes in Multihop IoT Networks using Dual Link Technologies and Unsupervised Learning", Open Journal of Internet of Things(OJIOT), Vol.4, Issue 1, PP.109-125, ISSN:2364-7108, 2018.
- [12] K. Somasundaram et al., "IoT-Attacks and Challenges", International Journal of Engineering and Technical Research (IJETR), Vol.8, Issue 9, PP.09-12, 2454-4698(Print), ISSN: 2321-0869(Online), September 2018.
- [13] Rupali Sachin Vairagade et al., "A Study of various authentication mechanisms towards the secure Internet of Things networks", Control and Cybernetics, Vol.49, No.4, PP.393-418, January 2021.

- 
- [14]Ms.K.Devipriya et al., "Enhancing Security in IoT Platform Using Secure Authentication Protocol", Turkish Journal of Computer and Mathematics Education, Vol.12, No.14, PP.5698-5708, 2021.
- [15]Gaurav Sharma,et al., "A Survey on Layer-Wise Security Attacks in IoT: Attacks, Countermeasures, and Open-Issues", Journal Electronics,PP.01-23,28 September 2021,Electronics 2021,10,2365.<https://doi.org/10.3390/electronics 10192365>.
- [16]Sher Ali et al., "Survey Paper on IoT Attacks and Its Prevention Mechanisms", Information Management and Computer Science (IMCS), Vol.3, Issue 2,PP.38-41,ISSN:2616-5961,doi:<http://doi.org/10.26480/imcs.02.2020.38.41>,28 December 2020.
- [17] AArul,et al., "A Review on Intrusion Detection Systems to Secure IoT Networks", International Journal of Computer Networks and Applications(IJCNA),Vol.9,Issue 1,PP.38-50,ISSN:2395-0455,doi:10.22247/ijcna/2022/211599,January- 28 February 2022.
- [18] Ali Hamid Farea et al., "Enhancement Trust Management in IoT to Detect ON-OFF Attacks with Cooja", International Journal of Multidisciplinary Studies and Innovative Technologies, Vol.5, No.2, PP.123-128, 2021.
- [19]Ansam Khraisat et al., "A Critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges", Cyber Security Springer Open Access, Vol.4, No.7, Issue 18, PP.01-27, ISSN: 42400-00077 <https://doi.org/10.1186/s42400-021-00077-7>,2021.
- [20] AKM Bahalul Haque, et al., "Security Threats and Research Challenges of IoT-A Review", Journal of Engineering Advancements, Vol.1, Issue 4, PP.170-182, 22 December 2020,<https://doi.org/10.38032/jea.2020.04.008>.
- [21]Anit Kumar et al., "Detection of Security Attacks on Edge Computing of IoT Devices through NS2 Simulation", 4th International Conference on Intelligent Circuits and Systems, PP.01-09, ISSN: 1742-6596,doi:10.1088/1742-6596/2327/1/012016.
- [22] Anca D.Jurcut, et al., "Introduction to IoT Security", ResearchGate, PP.01-37, 22 April 2020,doi:10.1002/9781119471509.w5GRef260.

## 3

## Various Approaches for Content Extraction from Web Pages based on Factors

**Deven M. Kene**

Department of Computer Science  
Vidyabharati Mahavidyalaya  
Amravati, Maharashtra  
Email: [dmk2006@rediffmail.com](mailto:dmk2006@rediffmail.com)

**Ather Iqbal**

Head,  
Department of Computer science  
Vidyabharati Mahavidyalaya  
Amravati, Maharashtra  
Email: [atheramt13@gmail.com](mailto:atheramt13@gmail.com)

### Abstract-

With the huge development of the internet and web publishing techniques generally create numerous information sources published as HTML pages on World Wide Web. So Extraction of Information has become an important task for discovering useful knowledge or information from the Web Pages. However, there is lot of redundant and irrelevant information on web pages. Navigation panels, Table of content (TOC), advertisements, copyright statements etc. A search engine and crawler system is one of the fundamental necessities of Information Extraction. Search engine takes into account only the informative content for indexing. With the informative content, web pages commonly have blocks that are not the main content blocks and are called the non-informative blocks or noise. Content in noise blocks will seriously harmful for information extraction, web mining, web searching. Noise affecting the precision of search and the size of index of search engines. So identifying the main content block is a key issue. In order to improve the performance of extraction of information, cleaning of Web pages becomes critical. There are various factors that you can consider in segmentation of web pages for Content extracrion from web pages. The main objective of this paper is to study various factors for web segmentation and content extraction. In the paper we proposed techniques, methods and factors for content extraction in web pages.

**Keywords-Search engine, information extraction, web content mining, web segmentation**

### I. INTRODUCTION

Web data mining tasks such as Web page clustering, classification, categorization, and information retrieval and information extraction [1]. Therefore these blocks are termed as the noisy blocks. Also, from the users' perspective only part of the information is useful for a particular application and the remaining information are noises. For improving the performance of traditional information extraction, it is necessary to differentiate valuable information from noisy content. Information contained in these noisy blocks can seriously hamper Web data mining task. Eliminating these noisy blocks is thus of great importance.

Now a day's World Wide Web (WWW) has become the main source of information for people, but the explosive growth of WWW has resulted in difficulties for individual user to process all these information. Search engines have been the various tools for users to find interested information on the web. Web contents such as multimedia data, structured i.e. XML documents, semi-structured i.e. HTML documents and unstructured data i.e. plain text [2] offer important information to the users and therefore be termed as informative contents. Other useful information on the Web is often accompanied by contents such as navigation bars, banner advertisements copyright notices etc [3] which can be termed as non informative contents.

In this paper, we present a various independent approach based on various factors for extraction of core contents from web pages. This approach extracts the data from each web page that are organised as well formatted XHTML documents. The different algorithms are

used based on various parameters for content extraction by filtering out the noise, and stores these contents into plain text form.

A web page usually contains various content such as contacts, information at the bottom, navigation bars, advertisements, or just some decoration components which are not related the topic of the web page. In information extraction, contents in these parts are all noise [4] information. Visitors to Web pages are only interested in the main content and have no use for the noisy content. Only the main content block of a web page contains the information we wants, we call this block *informative block*. So an accurate detection of informative block of a web page surely will improve the performance of information extraction. Content Extraction is a process of identification of what parts of a web page content the main textual content, thus ignoring the other irrelevant items of web page. Techniques belonging to the Web Content Mining such as classification and clustering, separation of block of web pages and removal of noisy blocks enable one to produce much better result for extracting useful information. Information retrieval or extraction applications consider all content on a Web page equally - e.g., with no differentiation between main and noisy content - there may be a decline in accuracy. It is necessary that such applications deal only with the main content of a Web page.

## II. RELATED WORK

Web sites are main source of information. The information contained on web sites is often mixed with non-informative content, is one of the main issues on the web. Eliminating the non-informative blocks will help in improving the process of information extraction. One approach to identify the informative blocks is web page segmentation. This section presents a summary of the techniques used to identify informative blocks within a web page.

Lin and Ho [5] proposed a method named info discoverer in which they used <table> tag to divide the web page into blocks. Then they extracted features from blocks and calculated entropy value of these features. Then this entropy value is used to determine whether the block is Informative or not.

Kao and Lin [7] proposed a method in which they used HITS (Hyperlink Induced Topic Search) algorithm to get a concise structure of web site by removing irrelevant structures. On the filtered structure they performed info discoverer method. This method is better than info discoverer because instead of using the whole web page they experimented on the filtered structure.

Today, the most frequently used segmentation algorithm is Vision-based Page Segmentation (VIPS), proposed by [2] from Microsoft Research. The vision-based method utilizes visual clues in a Web page. Chen proposed a method that considers visual information such as height, length of node zone and separation information. Yang proposed the VIPS (Vision-based Page Segmentation) algorithm by considering vision information and heuristic rules to identify blocks [14]. VIPS utilizes many visual cues such as element size, background color, font size and *etc.* to build a visual partition tree of a web page. Each node in the tree is a visual block of the web page. Kao [7] proposed WISDOM (Web Intrapage Informative Structure Mining Based on Document Object Model) method. This method evaluates the amount of information contained in node of DOM (Document Object Model) tree with the help of information theory. It first divides the original DOM tree into subtrees and chooses the candidate sub trees with the help of assigned threshold. Then a top-down and greedy algorithm is applied to select the informative blocks and a skeleton set which consist of set of candidate informative structures. Debnath [8] gave four algorithms content extractor, feature extractor, k-feature extractor and L-extractor for separating content blocks from irrelevant content. Content Extractor algorithm finds redundant blocks based on the occurrence of the same block across multiple Web pages. Feature Extractor algorithm identifies the content block with help of particular feature. K-



Feature Extractor, algorithm uses a K-means clustering which gets multiple blocks as compared to Feature Extractor that selects a single block. L-Extractor algorithm combines block-partitioning algorithm (VIPS –Vision based Page Segmentation algorithm) [1] with support vector machine to identify content blocks in a web page. Content-Extractor and Feature Extractor algorithms identifies primary content blocks by i) looking for blocks that do not occur a large number of times across web pages and ii) looking for blocks with desired features respectively. They identify primary content blocks with high precision and recall, reduce storage requirements for search engines, and result in smaller indexes. Performance evaluation shows that content extractor significantly outperforms the entropy based algorithm proposed by Lin and Ho in terms of accuracy and run-time. In content-extractor algorithm Debnath used the same basic concept used by Lin, that a <TABLE> tag is used to design maximum web pages. They make use of some other html tags also while designing the algorithm. Similar blocks across different web pages obtained from different web sites can also be identified using this algorithm.

Kang and Choi [10] proposed algorithm RIPB (Recognizing Informative Page Blocks) using visual block segmentation. This method also partitions web page into blocks based on VIPS. Similar structure blocks are grouped into clusters. A linear weighted function is applied to determine whether the block is informative or not. Uzun[11] proposed hybrid approach which combines automatic and manual techniques together for extraction process. Machine learning methods are used which draw rules for extraction process.

Wang [7] proposed a method which is based on fundamental information of web pages. This method extracts information from each web page and thereby combining that information to get site information. The information extracted is text node i.e. the data present in the tags, word length, menu sub tree which is a sub tree having text node length less than 5, menu item information, and menu instance information. Huang proposed a method employing block pre clustering technology. This method consists of two methods- matching phase and modeling phase. In matching phase, it first partitions the web page into blocks based on VIPS (Vision based Page Segmentation algorithm) [1]. Nearest neighbor clustering algorithm is used to cluster these partitioned blocks based on similar structures. Importance degree is associated with each cluster and clusters with importance degree are stored in clustered pattern database. In modeling phase, when a new web page comes it is first partitioned into blocks and then these blocks are matched with clustered pattern database to get the importance degree of these new partitioned blocks. Entropy evaluation is done on these blocks to know whether they are informative or not.

### **III. APPROACHES, METHODS AND FACTORS FOR WEB CONTENT EXTRACTION**

There are many methods proposed for the main content extraction. By using heuristic rules, determine whether an element of page is textual or not. Here we propose a wrappers language for extracting main content from web pages. The factors such as link density, Rules are considered. In this method the Vision based Page Segmentation algorithm considering for vision information and heuristic rules to identify blocks. [13]

A Novel approach for content extraction from web pages uses WLR. Word to leaf ration (WLR) combines with link attributes [12] of nodes for content extraction.

The popular Web page segmentation algorithms are DOM-based. In DOM-based segmentation, tag information is used to divide a Web page based on the Document Object Model (DOM). The DOM has a tree structure in which each node contains one of the components from an HTML tag. Web pages split using some relatively simple DOM nodes such as the <P>, <TABLE>, and <UL> nodes for further conversion or summarization.

We propose a web content extraction technique build on Entropy based Informative Content Density algorithm (EICD). The proposed EICD algorithm initially analyses higher text density content. Further, the entropy-based analysis is performed for selected features.

Here Text density, Content Ratio ,Page information density and Tag information density factors considering for content extraction.[14]

we proposed an algorithm for extracting the core contents of web pages using pattern matching approach that transforms the contents of web pages automatically in to plain text form. This approach deals with web pages of any size and extracts core contents with efficiency and high accuracy. The algorithm extracts high quality contents with efficiency and accuracy. Here we consider a Pattern as factors for content extraction.[15]

A simple but effective approach, named layout based detachment approach (LBDA). The proposed approach extracts the main content from the web page and removes the irrelevant information like header, footer contents, navigation bars, advertisements and other noisy images. This methodology uses tag tree parsing to get the analysis structure, block acquiring page segmentation method to remove unwanted tags, and data extraction to retrieve the necessary contents. It can eliminate noise and extract the main content blocks from web page effectively and display the essential content to the users. Here Time, Storage and accuracy factors consider for content extraction.[16]

The NEWSD (News Explorer for web Streaming Data) is a GUI text mining tool for the classification of web data is developed. In this tool feature extraction and classification apply on various news web sites for content extraction. Classifiers namely Naïve Bayes and J-48 are considering for results. Accuracy is main factors consider here.

#### IV.CONCLUSIONS

Informative Extraction of web content blocks from web pages is very important because web pages are unstructured and its number is growing at a very fast rate. Content Extraction is useful for the human users as they will get the required information in a time efficient manner.To extract the main content of a web page to prevent the treatment and processing of noisy, irrelevant and useless information is needed. We have presented some approaches for extracting main content from web pages and also studied a new approach for content extraction from web pages. In this paper we studied techniques for the extraction of content blocks based on the various factors for better and effective results. After implementation of various factors on Web page Segmentation and content extraction we get accurate and time efficient result.

#### REFERENCES

- [1] Lan Yi, Bing Liu, Xiaoli Li, "Eliminating Noisy Information in Web Pages for Data Mining," ACM SIGKDD, August- 2003.
- [2] XuhongZhang,Yanqing Zhang, and Frank "Vision based Web Page Block Segmentation and Informative Block Detection," International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 265- 269,IEEE Computer Society,2013.
- [3] S.H. Lin and J.M. Ho, "Discovering informative content blocks from web documents," Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ', pp. 588–593, 2002.
- [4] Christian Kohlschütter and Wolfgang Nejdl, "A Densitometric Approach to Web Page Segmentation," ACM SIGKDD, 2008.
- [5] S.Lin and J.Ho, "Discovering informative content blocks from web documents", Proceedings of ACM SIGKDD International Conference on Information Retrieval, pp. 450-453, 2002.

- 
- [6] Stevina Dias and Jayant Gadge, "Identifying Informative Web Content Blocks using Web Page Segmentation," *International Journal of Applied Information Systems (IJ AIS)*, Vol. 7, no. 1, pp.37-41, April- 2014
- [7] H.Kao and J.Ho, "WISDOM: web intrapage informative structure mining based on document object model," *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 5, pp. 614-627, 2005.
- [8] S.Debnath, P.Mitra, N.Pal and C.Giles, "Automatic identification of informative sections of web pages", *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 9, pp. 1233-1246, 2005.
- [9] D.Cai, S.Yu, J.-R.Wen and W.-Y.Ma, "Extracting content structure for web pages based on visual representation", *Proceedings of Fifth Asia Pacific Web Conference*, pp. 406-417, 2003
- [10] J.Kang and J.Choi, "Detecting informative web page blocks for efficient information extraction using visual block segmentation", *International Symposium on Information Technology Convergence*, pp. 306-310, 2007
- [11] E.Uzan, H.Agun and T.yerlikaya, "A hybrid approach for extracting informative content from web pages," *Information Processing and Management*, vol. 49, no. 4, pp. 1248-1257, 2013.
- [12] S.Shen and H.Zhang, "Block-level links based content extraction," *Fourth International Symposium on Parallel Architectures, Algorithms and Programming*, pp. 330-333, 2011.
- [13] Ahmad Pouranmini and Shahram Nasiri, "Web Content Extraction using Contextual Rules," *2<sup>nd</sup> International Conference on Knowledge-Based Engineering and innovation*, pp. 1014-1018, IEEE Computer Society, 2015.
- [14] Aanshi Bhardwaj and Veenu Mangat, "A Novel Approach for Content Extraction from Web Pages," *Proceedings of 2014 RAECS UIET Panjab University Chandigarh*, pp.978-980, IEEE Computer Society, 2014.
- [15] Sandeep Sirsat and Dr. Vinay Chavan, "Pattern Matching for Extraction of Core Contents from News Web Pages," *Second International Conference on Web Research (ICWR)*, pp. 13-18, IEEE Computer Society, 2016.
- [16] Dr. Anna Saro Vijendran C Deepa, "LBDA: A NOVEL FRAMEWORK FOR EXTRACTING CONTENT FROM WEB PAGES," *International Conference on Advanced Computing and Communication Systems (ICACCS -2013)*, Dec. 19 – 21, 2013, Coimbatore, INDIA.

## 4

## Overview and Classification of Social Security Attacks using Online Social Networking for Rumour Blocking

**Mrs.Shital M. Mohod**

Asst. Professor.  
Vidya Bharati mahavidyalaya,Amravati

**Prof. Ather Iqbal**

Head,Dept.Computer Science,  
VidyaBharati Mahavidyalaya,Amravati

### Abstract

The online social networking providers destine to secure their users; but the intruders and attackers are able to outsmart the security measures by exploiting user's privacy, identity and confidentiality using several techniques. An online social network(OSN) are permanent presence in today's personal and professional of a huge segment of the population. OSN also referred to as a virtual community is a website on the internet that serves as an ultimate location for people from different geometric locations to talk, share photos, ideas and interests, or make new friends. With the rapid increase in popularity and large number of user base, the online social networks also face an alarming rate of increase in security treats and rumors. On the other hand they become a cannel for the spreading of malicious rumors or misinformation. Most of the users in social networking sites might be unaware of the existence of these critical threats. We study different types of attacks to fight against rumors on social network.This paper highlights an overview and classification of Sybil, malware,DistributedDenial-of-service(DDOS),spam attacks.

**Keywords:** Attacks, OSN, DDOS, Social Security Network, Rumor Blocking

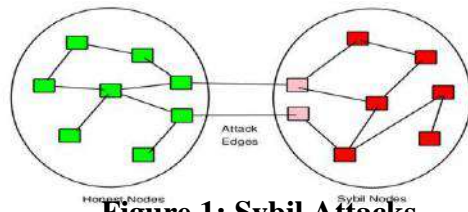
### 1. Introduction

With the increasing popularity of online social networks such as Twitter,Facebook,Renren and so forth. Rumors can spread farther, quicker and even with more terrible effect. An online social network have become a mainstream cultural phenomenon for millions of Internet users. In real world situation, Rumors exist in almost every domain of society. Example given, a Rumor generated in Twitter said that the president of Syria is dead, which hit the twitter greatly and was circulated fast among the population. We study different types of attacks to fight against rumours on social network. OSN services handle user's information and manage all user;s activities in the network. Being responsible for the correct functioning of its services and maintaining a profitable business model.Indirecting this translates into ensuring that their users continue to happily use their services without becoming victims of malicious actions.However attacks such as Sybil,Distributed Denile of service(DDOS) ,Spam and Malware. OSN's may translate into reputation damage.Service distrupction or other consequences with direct effect on OSN.

These attacks are aimed at the service provider itself by threatening its core business. These attacks can be performed by number of ways. However attacks on the OSN exploit the social graph of the OSN and victimised more users by propagating rapidly.So,the priority of the service provider would be identify and stop the propogation we brifly discuss different types of attacks on OSN's.

## 2. Types of Attacks

### 2.1 Sybil Attack:-



**Figure 1: Sybil Attacks**

In the Sybil attack, the malicious user claims multiple identities to compromise the whole network. Sybil attacks can be used to change the overall ranking in voting applications, access resources or to break the trust mechanism behind a P2P network. A P2P network is built on the assumption that each entity in the network holds a single identity. When an adversary introduces many bogus identities with a single entity or with no entity at all, a Sybil attack occurs. Using Sybil identities, an adversary may provide false opinions for his/her evil benefits, limit the amount of resources reaching each node, break the trust mechanism in a P2P network and may even cause a Denial-of-Service attack (DoS)[1]. In the initial researches to deal with Sybil attacks, network architectures were re-designed and secure mechanisms such as digital signatures and digital analyzers were used to mitigate the Sybil attacks[2]. Much effort has gone into the study of trust relationships in social networks [1][2][3][5] and community based schemes to reduce the influences of Sybil attacks [6][7].

#### 2.1.1 Classification of Sybil attacks

##### (i) Direct vs. In-Direct communication

the attacker must consider the type of communication between honest nodes and Sybil nodes [2][5]. If the communication between honest node and Sybil node is direct, i.e. if the attacker can directly communicate with the honest node using fake identities, it is a case of direct communication. However, if the attacker has to use his legitimate identity to communicate with the honest node, and then divert the Sybil data to the honest node via the legitimate node, it is the case of indirect communication. It is easier for the attackers to launch Sybil attacks in case of direct communication and it is also more difficult to detect such attacks.

**(ii) Busy vs. Idle:** In a P2P network, normally, only few Sybil identities participate in the network while the others remain idle. The power of the Sybil attacker comes from the number of identities he or she holds. If an attacker could afford to get fake identities easily, he or she can make the identities appear more realistic by making them leave and join the network multiple times pretending as an honest node. However, if the number of the Sybil identities are limited, the Sybil identities must participate simultaneously to launch an attack[5].

**(iii) Simultaneous vs. Non Simultaneous** A simultaneous attack can be performed by involving all the Sybil identities simultaneously or a single physical node can change its identities in regular time slots to appear like all the identities are involved simultaneously. In non-simultaneous attack, an attacker may bring all his identities into the network slowly over a period of time involving only few identities each time. This can be done by pretending that one identity is leaving the network while the other identity is joining the network. As honest identities generally tend to leave and join the network number of times, the malicious node won't be suspected if they pretend to leave or join the network now and then using different identities[8].

**(iv) Insider vs. Outsider** The impact of the Sybil attack depends on whether the attacker is inside or outside the distributed network. If the adversary is part of the network and holds at

least one real identity, then the attacker is called an Insider, otherwise he or she is an outsider. An insider may introduce many fake identities, and pretend to communicate with other nodes using his fake identities. However, for an outsider, it is difficult to introduce Sybil identities into the network,

## 2.2 Distributed Denial-of-service:

A distributed denial-of-service (DDoS) attack is an attack in which multiple compromised computer systems attack a target, such as a server, web site or other network resource, and cause a denial of service for users of the targeted resource. The flood of incoming messages, connection requests or malformed packets to the target system forces it to slow down. Exploited machines can include computers and other network resources such as IoT devices. In a typical DDoS attack, the assailant begins by exploiting a vulnerability in one computer system and making it the DDoS master. The attack master system identifies other vulnerable systems and gains control over them by either infecting the systems with malware or through bypassing the authentication controls (i.e., guessing the default password on a widely used system or device). A computer or networked device under the control of an intruder is known as a zombie, or bot. The attacker creates what is called a command-and-control server to command the network of bots, also called a botnet[10]. The person in control of a botnet is sometimes referred to as the botmaster (that term has also historically been used to refer to the first system "recruited" into a botnet because it is used to control the spread and activity of other systems in the botnet). Botnets can be comprised of almost any number of bots; botnets with tens or hundreds of thousands of nodes have become increasingly common, and there may not be an upper limit to their size.

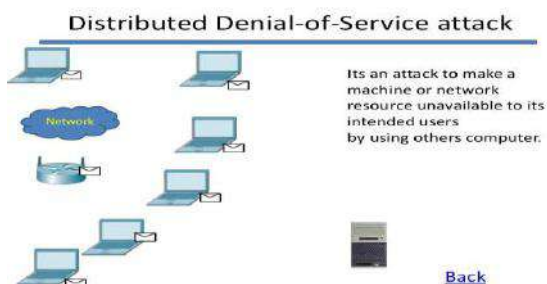


Figure 2: DDOS

### 2.2.1 Classification of DDoS attacks

There are three types of DDoS attacks. Network-centric or volumetric attacks overload a targeted resource by consuming available bandwidth with packet floods. Protocol attacks target network layer or transport layer protocols using flaws in the protocols to overwhelm targeted resources. And application layer attacks overload application services or databases with a high volume of application calls. The inundation of packets at the target causes a denial of service. While the things comprising the internet of things (IoT) may be useful to legitimate users, in some cases, they are even more helpful to DDoS attackers. The devices connected to IoT include any appliance into which some computing and networking capacity has been built, and, all too often, these devices are not designed with security in mind[9]. Devices connected to the IoT expose large attack surfaces and display minimal attention to security best practices. For example, devices are often shipped with hard-coded authentication credentials for system administration, making it simple for attackers to log in to the devices. In some cases, the authentication credentials cannot be changed. Devices also often ship without the capability to upgrade or patch device software, further exposing them to attacks that leverage well-known vulnerabilities. Internet of things botnets are increasingly being used to wage massive DDoS attacks.

## 2.3 Spam Attack

is an endless repetition of worthless text or image. Spam can spread out in any information systems like emails, web, social network sites, and blogs or in review platforms. The concept of web spam was introduced in 1996 [11] and it soon become key challenges for search engine industry [12]. Nowadays the major search engine companies have identified adversarial information retrieval [13] as top priority because of multiple negative effects caused by spam, and also the appearance of new challenges in the field of research. First spam spoils the quality of research and prevents the legitimate websites of revenue that might earn in the absence of spam. Second it weakens the trust of user in a search engine provider which is a notable issue since the user can easily continue his search form one search engine to other. Spam refers to the use of electronic messaging systems to send out unrequested or unwanted messages in bulk.



Figure 3:Spam Attacks

### 2.3.1 Classification of Spam Attack

**i) Social network spam:** In past few years the development of social networking sites is very high. The people communicate with their friends and chat or share multimedia contents with them. Sites like facebook, twitter are constantly among top 20 most viewed websites on the internet [13]. People spent more time on social network compared with other sites. The increase in popularity of social networks allows them to collect a huge amount of personal information about the users, their friends, habits and also their wealth information. In social network a person can reach any person which is attracted by the malicious parties.. As for Twitter, [12] ran an experiment on Twitter spam. Regarding the drawbacks in Bayesian spam filter an user-friendly spam filter called Social network Aided Personalized and effective spam filter (SOAP) is used.. social closeness spam filtering, social interest based spam filtering, and adaptive trust management.

**ii) Email spam:** The most common communication in the internet is using email communication. With the vast growth in email and its popularity unsolicited e-mail (spam) also emerged very quickly with almost 90% of all email messages. i.e., over 120 billion of these messages are sent each day [12]. The cost of sending these e-mails is very close to zero being easy to reach a high number of potential consumers [13]. In this context, spam consumes resources; time spent reading unwanted messages, bandwidth, CPU, disk, being also used to spread malicious content. The email system design can easily be exploited by spammers who send inaccurate information. All email on the Internet is sent via a protocol called Simple Mail Transfer Protocol ("SMTP").SMTP is designed to capture information about the route that an email message travels from its sender to its recipient. In actuality, the SMTP protocol provides no security, email is not private, it can be altered en route, and there is no way to validate the identity of the email source.

**iii) Image spam :**Recently, spammers have proliferated "image spam", emails which contain the text of the spam message in a human readable image instead of the message body. It consists in embedding the spam message into images which are sent as email attachments. Its

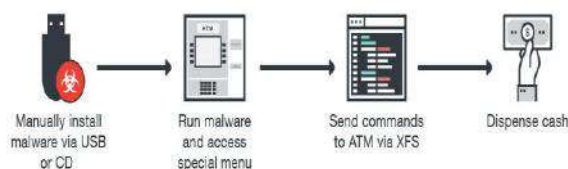
goal is to circumvent the Analysis of the emails' textual content performed by spam filters, including automatic text classifiers. Since attached images are displayed by default by most email clients, the message is directly conveyed to the user as soon as the email is opened. The simplest kind of image spam can be viewed as a screen shot of a plain text written using a standard text editor

**iv) Click spam.** Here the spammers generate fraud clicks and make the control function towards their websites. To achieve the goal spammers submit queries to search engine and click on the links point to the target pages [12, 13]. Online advertising is other incentive for spammers to generate fraudulent clicks [13]

**V) Content Spamming :** Content spamming involves changing the logical view that the search engine has over the page contents. An example of content spamming is keyword stuffing which involves placement of keywords within the webpage to raise the keyword count.

### Malware Attack

Malware does the damage after it is implanted or introduced in some way into a target's computer and can take the form of executable code, Script, active content and other software[14]. The code is described as computer viruses, worms, Trojan Horses, Spyware[15]. The term malware comes from combining the two words malicious and software, and to be used to indicate any unwanted software. any code added, changed, or removed from a software system.[16].The purpose of Malware is to cause damage or penetrate users computer for the purpose of hacking personal data for illegal activity such as financial crimes. Many DoS viruses, and the Windows Explore Zip worm, are designed to demolish files on a hard disk, or to corrupt the file system by writing void data to them.



**Figure 4:Malware Attacks**

#### 2.4.1 Classification of Malware Attack

Several malware classifications have been issued so far, depending on some of their characteristics. The purpose of such classifications is to facilitate the tracking of authorship, correlating information, identifying new variants [17]. However, The classification made, is to categorize the major common malware types into groups depending on the network and web usage.

##### i) Network-based Malware

Spyware is a kind of malware that is installed secretly on a user computer for the purpose of collecting information about users without their knowledge [17]. Even reputable vendors of software like Microsoft and Google, intentionally, collect information of their users using spywares [18].

Cookies are some information stored on user's computer by their web browsers. The main purpose of cookie is to authenticate users depending on the information stored in, storing site preferences and server-based session.



Trojan horse is a code that appears to be a useful program, but actually it steals information or corrupts data [17,18].

Botnet is a collection of infected computers (contains bot software embedded in it) that have been taken over by hacker and used to perform malicious functions, without the hackers having to log into the client's computer. Botnet can make DoS attack as many clients' bots, under control of hacker bot, having a role of attack [19, 20].

## ii) Ordinary Malware

Virus is any software code that has the ability to replicate itself, during infection, into any other application software or a document. Viruses can do harmful functions on a user machine; it can make destruction to the whole system from infected device to uninfected one [18, 19, 20]. Worm is any software code that has the ability of self replicating on victim computer. Worms are independent; they don't need for a host program to start lifecycle[20].

## 3. Comparative Analysis of Online Social Security Attacks

Table 1: Comparison of Social Security Attacks:

Attacks Type	Sybil	DDoS	Spam	Malware
Layers	Network	Transport,Application	Application	Network
Techniques	Light Weight Sybil Attack Detection	Defence techniques	Rulr Based scoring system	Signature and Detection
Methods	Robust,Lightweig ht	Artificial Neural Network	Spammers	Pre-pending, Embedding,Post-pending
Virus Activated	Worm	Botnet	Worms	Trojan Horse
Mode of Attacks	Rumor(Bogus Identities)	Intruder	Repetition of worthless text or image	Intrusive Code
Protocol	P 2 P	UDP,TCP/IP,HTTP, SMURF	SMTP,VOIP	UDP,HTTP,SOCKS 4/5
Advantages	i)Efficient in large overhead ii)No clock synchronization	i) Detecting and stopping a DDoS attack at the Source providesr.  ii) minimum damage is done on legitimate traffic .	i) it is essential to send as many messages as possible in a short period of time. ii) the transitional period as only one service needs to be maintained instead of two parallel running services	i) Protection from Phishing Attacks ii) Provides Robust Web Protection
Disadvantages	i) reliability is lesser ii) May encourage attackers economically	i)Detecting DDoS attacks at source end is difficult because sources are widely distributed across the network and a single source behaves like a normal traffic. ii)The difficulty of deploying system at the source end	i) Limiting the number of emails that can be send ii) Analyzing the messages to determine if they contain spam or no	i) Security defects in software ii)_Insecure design or user error

#### 4. Conclusion:

In this paper, we study how social network security attacks occurs to fight against rumors on social network and their classification. In Sybil attack, an insider may introduce many fake identities and pretend to communicate with other nodes using his fake identities and for an outsider it is difficult to introduce sybil identities into the network. In DDoS, multiple compromised computer systems attacks a target such as server, website or other network resources and exploiting a vulnerability in one computer system and making it the DDoS master. Using SMTP, VoIP protocols, Spam is essential to send many messages in short period of time. There is an endless repetition of worthless text or images. spam can spread out in any information system like E-mails, Web, Social Network Sites. using malware attack any code added, changed or removed from a software system in order to intentionally cause harm or disturb the intended function of the system that encompasses viruses, Trojans and other intrusive code. The purpose of such classification is to facilitate the tracking of authorship, correlating information, identifying new variants.

#### References

- [1] H. Yu, M. Kaminsky, P.B. Gibbons, and A. Flaxman. Sybil Guard: Defending against Sybil attacks via social networks. In SIGCOMM, 2006.
- [2] H. Yu, M. Kaminsky, P.B. Gibbons, and A. Flaxman. Sybil Limit: A near optimal social network defense against Sybil attacks. In IEEE Symposium on Security and Privacy, 2008.
- [3] J. Newsome, E. Shi, D. Song, A. Perrig. The Sybil Attack in Sensor Networks: Analysis & Defenses, In Proc. of ACM IPSN, 2004.
- [4] W. Chang and J. Wu. A Survey of Sybil Attacks in Networks. In publications of computer and Information Sciences, Temple University, Philadelphia, 2013
- [5] G.V. Rakesh, S. Rangaswamy, V. Hegde, G. Shoba. A Survey of techniques to defend against Sybil attacks in Social Networks, In IJARSCCE, 2014.
- [6] W. Wei, F. Xu, C.C. Tan, Q. Li. SybilDefender: Defend Against Sybil Attacks in Large Social Networks, In Proc of IEEE INFOCOM, 2012.
- [7] L. Shi, S. Yu, W. Lou, Y. T. Hou. SybilShield: An agent-Aided Social Network-Based Sybil Defense among Multiple Communities, In Proc of IEEE INFOCOM, 2013
- [8] Manju V C "Sybil attack prevention in Wireless Sensor Network", IJCNWMC 2014.
- [9] A. Harrison, "The denial-of-service aftermath," Feb. 2000, <http://www.cnn.com/2000/TECH/computing/02/14/dos.aftermath.idg/index.html>.
- [10] [www.academia.edu/2963956/A\\_Brief\\_Review\\_of\\_Denial-of-Service\\_Research\\_Papers](http://www.academia.edu/2963956/A_Brief_Review_of_Denial-of-Service_Research_Papers)
- [11] Wikipedia, "Spam" [http://en.wikipedia.org/wiki/Spam\\_\(electronic\)](http://en.wikipedia.org/wiki/Spam_(electronic))
- [12] <https://www.ijcaonline.org/archives/volume157/.../vairagade-2017-ijca-912633.pdf>
- [13] <https://ieeexplore.ieee.org/abstract/document/7226077/>
- [14] <https://ieeexplore.ieee.org/document/8094101/> Vinod, P., et al., Survey on Malware Detection Methods. 2009
- [15] McGraw, G. and G. Morrisett, Attacking Malicious Code: A Report to the Infosec Research Council. IEEE Softw., 2000. 17(5): p. 33-41.
- [16] Xufang, L., P.K.K. Loh, and F. Tan. Mechanisms of Polymorphic and Metamorphic Viruses. in Intelligence and Security In
- [17] Egele, M., et al., A survey on automated dynamic malware-analysis techniques and tools. ACM Comput. Surv., 2008. 44(2): p. 1-42.
- [18] Idika, N. and A.P. Mathur., A Survey of Malware Detection Techniques. 2007.
- [19] Nataraj, L., Karthikeyan, S., Jacob, G. and Manjunath, B. (2011) Malware Images: Visualization and Automatic Classification. Proceedings of the 8th International Symposium on Visualization for Cyber Security, Article No. 4.
- [20] Nataraj, L., Yegneswaran, V., Porras, P. and Zhang, J. (2011) A Comparative Assessment of Malware Classification Using Binary Texture Analysis and Dynamic Analysis. Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence,

## Big Data: Security and Security Challenges

**Miss .Dipika S. Harode**

Department of Computer Science,  
Vidyabharati Mahavidyalaya, Amravati  
[dipikaharode11@gmail.com](mailto:dipikaharode11@gmail.com)

### Abstract

Before 15 -20 years data was too limited because use of Social media, Online Transactions, E-Commerce, etc. was not in that extent, it was easy to store, process and protect the data. The amount of data in world is growing day by day. Data is growing because of use of internet, smart phone and social network. But big data is a double-edged sword. It brings convenience to people and brings certain risks. In the process of data collection, storage, and use, it can easily lead to the leakage of personal information, and the fact that data is difficult to discern. if data is not well protected from threats like phishing, hacking etc. all these processing becomes futile as if data falls in wrong hands, it could be misused. There are many ways to maintain data security and privacy but still it could be violated if not carried out properly. So while dealing with data Security and its challenges becomes prime concern in order to protect it from attacks. How to ensure big data security and what security challenges we are facing has become one of the hot issues in the current stage of research. This paper discusses the concept of big data and surveys the current research carried out on security as well as security challenges in big data. . The security factors and security challenges are discussed.

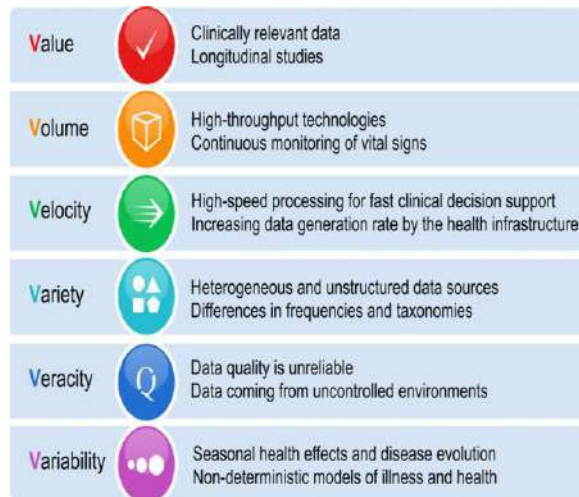
**Keywords** *Big data, phishing, hacking, data security. Data security*

### Introduction

With the rapidly increasing amounts of data produced worldwide, the amount of data to be processed continues to witness a quick increase. We all are living in the era where data is generating enormously, even the data generated in last few years is more than the data generated earlier in the whole century and it is clear that future is filled with more data , So that means we are going to deal with a huge amount of data which keep on increasing. This data is generating from various platforms like Social media, E- commerce websites, Stock market, data of particular organizations and list goes on. This data holds valuable information like how users share, view or engage with content in case of Social media; ratings and reviews of customers, preferences, shopping behavior and payment information.

In every organization, data is most important asset, not only for computer based industries but also for other organizations such as governments, healthcare, education, engineering and technology sector, manufacturing, and retail sector. A human is a social person always living in a society and interacting daily with each other. Due to technology advancement and applications, such as smart mobile devices, sensors Internet of Things, cyber-physical systems, social networks, YouTube, charting software etc produces a huge amount of data daily. data, so new term have come to be known as Big Data which process a huge amount of these types of data. Big data analytics are used every day by organizations like stock market and retail, etc. to improve business. New technology comes with new issues. The term big data defined as "a new generation of technologies and architectures, designed to economically separate value

from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and analysis". Still the Big Data challenges are not only define in 6V's such as **volume, velocity, variety, veracity, value**, variability, venue, vocabulary, and vagueness but also others factor like data quality, data privacy and data security too play an important role.



**Fig1. Six V's in Big Data**

In recent years, various mechanisms have been developed to ensure big data security and security challenges. This security and security challenges issue cannot be ignored and it is necessary to protect the sensitive personal information of customers from online criminals because data is an important asset to any organization and also for the millions of customers who trust such organization with their information.

These mechanisms can be categorized on the base of big data life cycles, such as data generation, storage, and processing. In data generation phase, access restrictions and falsifying data techniques are used for data privacy technique This gives criminals an opportunity to collect information on the Internet and then conduct illegal activities such as reselling, fraud, etc., not only for people. Life has brought troubles and brought economic losses, which has seriously affected social stability and harmony. In the era of big data, how to deal with security and privacy issues in the context of big data is an urgent need for people to have a good solution.

### **Big data security**

It refers to the measures and practices implemented to protect large volumes of data against unauthorized access, breaches, and malicious activities. Securing big data involves 3 main phases:

- 1 Ensuring the safe transfer of data from source locations, typically in the cloud, for storage or real-time ingestion.
2. Safeguarding data within the storage layers of the big data pipeline.
3. Maintaining the privacy of output data, including reports and dashboards, which contain insights obtained from data analysis using tools like Apache Spark.

Big data security refers to the security measures and mechanisms implemented within a big data environment to protect the data, infrastructure, and applications involved in big data processing. It focuses on the technical aspects of securing the various components of a big data

ecosystem, including data storage systems (e.g., Hadoop clusters, data warehouses), data processing engines, data pipelines, and the data itself. Big data security includes data encryption, access controls, authentication, authorization, monitoring, threat detection, and data masking specific to the big data environment.

### Essential Methods for ensuring big data security:



*Fig 2. Data security controls*

### Encryption:

It assumes a critical role in this endeavor. The imperative is to establish scalable encryption practices encompassing data at rest and data in transit within the comprehensive Big Data pipeline. Scalability takes precedence here as data encryption should extend its protective reach to encompass various analytics tools, their outputs, and storage formats like No SQL. Encryption's potency emerges from its capacity to render data indecipherable, even when malicious actors intercept data packets or gain access to sensitive files.

### User access control

Effective access control is vital to tackling big data security issues like insider threats and excessive privileges. Role-based access management is a valuable method for overseeing access throughout various layers of big data pipelines. Following the principle of least privilege helps restrict access to only the necessary tools and data for a user's tasks.

### Cloud security monitoring

Due to the substantial need for storage and processing in big data workloads, cloud computing has become a practical choice for many enterprises. At the same time, vulnerabilities like exposed API keys and misconfigurations in cloud environments can't be ignored.

### Centralized key management

In a complex big data ecosystem, encryption security requires a centralized key management approach to ensure effective and policy-driven handling of encryption keys. Centralized key management also controls key governance from creation to key rotation. For businesses running big data workloads in the cloud, Bring Your Own Key (BYOK) is probably the best option that allows for centralized key management without handing over control of encryption key creation and management to a third-party cloud provider.

## Network traffic analysis

In a big data pipeline, a stream of data is continuously ingested from various origins, encompassing sources like real-time data from social media platforms and information from user endpoints. The analysis of network traffic serves as a means to gain insight into this traffic and identify any irregularities, such as the presence of potentially harmful data from IoT devices or the utilization of unsecured communication protocols.

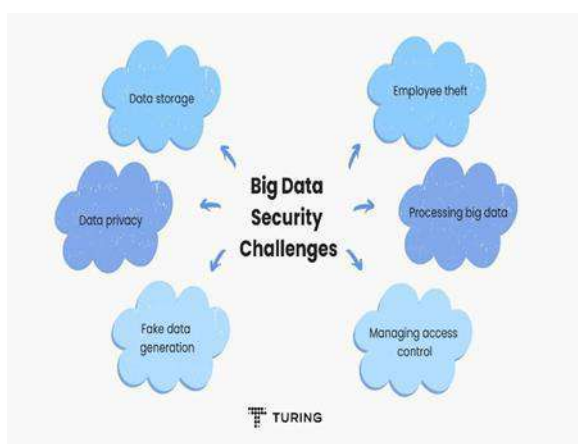
## Threat hunting

Threat hunting is a proactive effort to uncover hidden threats in your network. Led by an experienced cyber security analyst who uses real-world attack data and insights from security tools, its goal is to formulate hypotheses about potential threats. Big data can assist in this process by uncovering concealed insights within large sets of security data. For big data security enhancement, threat hunting involves examining datasets and infrastructure for signs of compromise in your big data environment.

## Data exfiltration detection

Security leaders worry about unauthorized data transfers in big data pipelines, where vast amounts of sensitive assets can be copied. Detecting data exfiltration requires monitoring outbound traffic, IP addresses, and network activity. Prevention involves tools for code security, misconfiguration checks, data loss prevention, and next-gen firewalls. Educating and raising awareness within your organization is essential.

## Big data security challenges



*Fig 3. Big data security challenges*

The continuously growing volume of data offers advantages and drawbacks. Enhanced data analysis can lead to better decision-making for businesses, but it also introduces security concerns, especially when handling sensitive information.

Here are some of the challenges in big data security that organizations need to address

## **Data storage**

Big data involves storing and processing vast amounts of data; securing it can be challenging. Big data systems store various data types, including unstructured, structured, and semi-structured data, making it difficult to implement security measures effectively for all data types. Moreover, data redundancy and replication are common in big data architecture, meaning that sensitive data may exist in multiple locations, which increases the risk of unauthorized access.

## **Fake data**

Fake data generation is another big data security challenge because it can be used to manipulate and deceive big data systems. This challenge can lead to inaccurate results and insights, forcing businesses to make wrong decisions. For example, criminals may generate fake product reviews to manipulate potential customers' purchase decisions. Besides that, fake data can be used to mask real data, making it easier for attackers to steal sensitive data.

## **Data privacy**

Data privacy is a significant challenge for big data security because big data systems often collect and store large amounts of personal data. It collects data from multiple sources, including online and offline activities, making it difficult for businesses to secure and maintain data privacy. Furthermore, big data systems involve sharing data with third-party applications and services that can increase the risk of data breaches and unauthorized access.

## **Data management**

A security breach can have severe repercussions, including the exposure of critical business information within a compromised database. To ensure data security, deploying highly secure databases with various access controls is essential. Robust data management systems offer extensive security measures, including data encryption, segmentation, partitioning, secure data transfer, and trusted server implementation.

## **Data access control**

Big data systems are highly complex and distributed, spreading data across multiple storage locations and servers. This makes it difficult to implement and manage access controls that can work for all data formats. Big data systems also store large volumes of data and share them with third-party applications and services. Managing access to such massive and diversified data is a major challenge, and the risk of unauthorized access to the data is always higher.

## **Data poisoning**

ML solutions, like Chabot, continuously improve through interaction with vast datasets, but this progress can be exploited through data poisoning attacks. This tampering with training data can compromise the model's ability to make accurate predictions, resulting in logic corruption, data manipulation, and data injection. Detecting outliers is a powerful defense against such attacks, helping separate injected elements from the existing data distribution.

## Employee theft

Every employee in an organization has some amount of access to the data, especially those who are involved in big data analysis. Some employees even have insider knowledge of the organization's data systems, including access controls, passwords, and security protocols. An employee with access to a big data system can exploit the authority to gain unauthorized access to sensitive data. They can also manipulate data to cause financial and reputational harm to the organization.

## Conclusion

In this paper we come to know that no matter how advanced big data technology is, the very first priority is securing and facing security challenges of big data in order to protect data from malicious attacks, ensuring safety of data to stop it from falling into wrong hands. And there are many techniques which can be used to protect big data from such harms like we have gone through De-Identification, Encryption. There are other technologies also like Data Cryptography, End point filtration etc. which are used to deal with Data security and privacy using different ways and algorithms. But even after having so many techniques to secure the data, there are few shortcomings of these techniques too as the amount, source, type and speed at which the data is generating, it is quite difficult to protect it from attacks .So we need more advance techniques to deal with Big Data Privacy and Security to put an end to such frauds.

## References

- [1] Big data security book by Shibakali Gupta, Indradip Banerjee, Siddhartha Bhattacharyya
- [2] Jayesh Surana, Akshay Khandelwal, Avani Kothari, Himanshi Solanki, Meenal Sankhla, Big Data Privacy Methods 2017 IJEDR, Volume 5, Issue 2, ISSN: 2321-9939.
- [3] Priyank Jain, Manasi Gyanchandani & Nilay Khare, Big data privacy: a technological perspective and review, 2016.
- [4] Alex Bekker, Buried under big data: security issues, challenges, concerns, Head of Data Analytics Department, Science Soft 2018.
- [5] M. Manikandakumar (Thiagarajar College of Engineering, India) and E. Ramanujan
- [6] <https://www.ijert.org/big-data-security-and-privacy>
- [7] <https://www.businesstechweekly.com/operational-efficiency/data-management/big-data-privacy-and-security-challenges/>
- [8] <file:///C:/Users/Lenovo/Downloads/25904185.pdf>
- [9] <https://www.businesstechweekly.com/operational-efficiency/data-management/big-data-privacy-and-security-challenges/#What-is-Big-Data-Privacy-and-Security>
- [10] <file:///H:/SecurityandPrivacyforBigData.pdf>
- [11] <https://maddevs.io/blog/big-data-security-best-practices/>
- [12] <https://www.google.com/search?q=methods+of+big+data+security+images&client=firefox-b->
- [13] [https://www.youtube.com/watch?v=w5gcoYq\\_3Cw](https://www.youtube.com/watch?v=w5gcoYq_3Cw)
- [14] <https://www.youtube.com/watch?v=8B2r4J7OPqc>
- [15] [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwich8ig5euDAX3a2wGHcB8BtsQFnoECAoQAQ&url=https%3A%2F%2Fwww.researchgate.net%2Fpublication%2F296702778\\_Big\\_Data\\_Security&usg=AOvVaw1urTulXBbCAh](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwich8ig5euDAX3a2wGHcB8BtsQFnoECAoQAQ&url=https%3A%2F%2Fwww.researchgate.net%2Fpublication%2F296702778_Big_Data_Security&usg=AOvVaw1urTulXBbCAh)



---

# Challenges in Devanagari Script-based CAPTCHA: A Comprehensive Analysis

**Anita B. Dube**

Shri Shivaji College of Arts, Commerce & Science, Akola, MH, India

## **Abstract:**

As the number of online services continues to grow and the threat of automated bots continues to increase, the need for strong security measures has never been more important. CAPTCHA, which stands for Completely Automatic Public Turing Test to Tell Computers Apart, is one of the most popular ways to distinguish between people and machines. A variety of CAPTCHAs are there, but they are in English. In a multilingual country like India, there is a need to develop Captcha in the native language. Devanagari is a script used to write Sanskrit, Hindi, Marathi, and many other Indian languages. This paper gives a comprehensive analysis of challenges that have been faced in setting up a CAPTCHA system using the Devanagari script. The purpose of this paper is to look at the weaknesses and issues that have been encountered in developing a reliable and robust CAPTCHA system based on this script.

**Keywords:** Captcha, Bots, Devanagari script

## **1. INTRODUCTION**

In recent years, internet applications such as email, social networks, blogs, and e-government sites have become a necessity for everyone. As the Internet continues to grow, so does the need for Security. CAPTCHA stands for "Completely automated public Turing test to distinguish Computers from Humans". It's a security feature used to distinguish human users from automated bots. CAPTCHA was created in the late 90s as a solution to the growing problem of automated scripts and bot attacks on the Internet. CAPTCHA's main purpose is to prevent unauthorized access to an online system, website, or application from spamming, fraud, and other unauthorized activities. A successful CAPTCHA meets the following criteria: Automatic generation of the test Quick and easy response to the test accepted by all human beings and resistance to attacks with a publicly known protocol. The Devanagari language is one of India's official languages and is spoken by the majority of people. Devnagri script is the base of most Indian languages, which includes Marathi, Hindi, Bengali, Gujranwala, Konkani, and other northern Indian languages[1][2].

### **1.1 Origin & Purpose**

A captcha is a security feature that is used on websites to differentiate between automated bots and humans. While CAPTCHA was first developed by John Lanford at Carnegie Mellon University[3], it was Mori Naor[4] who first introduced the idea of the Turing Test to differentiate between a human and a robot in 1996. The term "Turing test" in CAPTCHA refers to the test proposed by Alan Turing to determine a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. OCR can compromise CAPTCHA security. There are many different types of CAPTCHA, each with its benefits and drawbacks. CAPTCHA is a challenge-response test that is used to identify whether the user is a human being or not. The CAPTCHA is based on the reverse Turing Test. The Turing Test is

a test that can be used to identify whether a computer can understand a text, an image, a video, or an audio. The computer cannot understand a distorted text or a number.

Web services use Captcha for the following purposes:

- **Spam Prevention:** One of CAPTCHA's main goals is to stop automated bots from sending forms on websites. CAPTCHA does this by creating challenges that humans can easily solve but machines can't. This helps reduce the number of automated spam submissions[5].
- **Account Security:** CAPTCHA, is often used during account creation to verify that the new account is created by real human users and not automated scripts, which can be used to create thousands of fake accounts in a short period [6].
- **Security in Online Transactions:** CAPTCHA is used in online transactions and financial transactions as an extra layer of security. CAPTCHA ensures that human users initiate these processes to protect sensitive information and avoid automated attacks[7].
- **Data Protection:** CAPTCHA is used by websites and online platforms to protect against web scraping[7], which is the practice of using automated bots to collect large amounts of information from websites. The CAPTCHA challenges prevent these automated data harvesting attempts.
- **Protecting Website Registration:** It is used to protect several free email services (Yahoo, Gmail, Hotmail) from bot programs that register thousands of email accounts per minute using an automated script[7][8].
- **Bot Protection:** A CAPTCHA prevents automated bots from doing things like creating duplicate accounts, filling out forms, or doing things that could damage a website or an online service[7][8].

## 1.2 Early Captcha Design

There are four main types of CAPTCHA, which are: Text Based, Image Based, Audio Based & Video Based. Now a variety of Captchas are introduced. CAPTCHAs are AI problems that can't be solved by today's software or bots but can be easily solved by people. A client who gives the right answer to a question is considered a human otherwise, a bot.

The earliest CAPTCHA was based on distorted text that was hard for OCR (Optical Character Recognition) software to read. Users had to manually enter the characters from the distorted image to show they were human. This is the most common type of CAPTCHA. Google, Yahoo, and Microsoft have all had their own text-based CAPTCHA deployed for years. [9][10][11].CAPTCHA based on images has been suggested as an alternative to text media. The users are asked to perform an image recognition task[12][13]. Audio Captcha is based on the user's ability to recognize sound or speech. It was introduced as an accessible option for people who can't use the more popular visual CAPTCHA and for the visually impaired. The video-based CAPTCHA was created primarily to make CAPTCHA sessions engaging and creative [14]. A moving object is presented to the user and the user is asked to complete a task.

## 2. MOTIVATION OF DEVANAGARI SCRIPT-BASED CAPTCHA

Using the Devanagari script-based Captchas can serve many purposes, each one with its own motivation[15]

**Language Specificity:** In the captchas of any website or application, the security feature should be based on the language of the target user. The majority of these languages are Hindi language, Marathi language, Sanskrit language, etc. Therefore, if a website or application caters to users who mainly communicate in their local language, then one should use the script in the captchas[16].

**Cultural Sensitivity:** For web applications that emphasize cultural sensitivity, the use of Devanagari in captchas conveys a sense of respect for and acceptance of the language and culture of the target audience.

**Localization:** if a website is localized for areas where Devanagari scripts are widely used, including Devanagari in captchas is in line with the localization strategy. This can help users to feel more secure and understand the security features[16].

## 3. CHALLENGES IN DEVANAGARI SCRIPT-BASED CAPTCHA

As OCR became more advanced, the types of CAPTCHA tasks developed to include image recognition, image selection, mathematical problems, and distorted characters against a complex background.

**3.1 Linguistic Complexity:** Language complexity is a key design and implementation challenge for CAPTCHA. Language complexity refers to tasks that are simple for humans to solve, but challenging for automated scripts, especially those using machine learning algorithms. When it comes to linguistic complexity, CAPTCHA design challenges often center around using natural language elements. Below are some of the specific aspects of language complexity that challenge CAPTCHA design [17, 18]:

**3.1.1 Multiple character variations:** Languages often have different types of characters, including uppercase letters, lowercase letters, accents, and letterforms. In Devanagari, each character represents a consonant with an inherent vowel sound, and additional vowel sounds can be added. The use of diacritic marks can result in the addition of additional vowel sounds.

**3.1.2 Ligatures and conjunct characters:** Complexity is added to some languages, such as Devanagari or Arabic, or to scripts that have ligatures. However, automated systems may not be able to identify and isolate ligatures, making it difficult to create CAPTCHA that contains such language features.

**3.2 Font Variability:** Devanagari font variability is one of the most important challenges when designing a CAPTCHA based on the script. Devanagari script is a very complex script. It has a lot of different characters, ligatures, and variations. When designing a script-based CAPTCHA, font variability can be a problem for both human users and automatic recognition systems. Here are a few aspects of font variability in Devanagari scripts CAPTCHA:

**3.2.1 Font styles and shapes:** The Devanagari characters are written in different font styles with different shapes and forms. Using different font styles for CAPTCHA

characters can be a problem for automatic systems that can't generalize across different representations.

**3.2.2 Legibility concerns:** Some fonts may contain characters with complicated shapes or complex details, which can cause readability problems for both human and automatic systems. It is important to strike a balance between keeping characters legible for human users while introducing complexity for automatic recognition.

**3.3 Cultural and Regional Sensitivity:** Devanagari scripts are used not only in Hindi but also in several other Indian languages. The usage of the script varies from one region to another. Therefore, it is essential to take into account cultural and regional sensitivities when designing a CAPTCHA based on the script. It is also important to ensure that the CAPTCHA is effective, inclusive, and respectful of the language and cultural diversity of the country. Here are some things to keep in mind when designing a CAPTCHA based on script. When designing a script-based CAPTCHA, there are some factors to consider.

**3.3.1 User familiarity with script:** Different regions of the world may have their unique dialects of Devanagari. Therefore, it is important to consider how familiar users are with specific characters, ligatures, and writing styles for a good user experience.

**3.3.2 Regional variations in script usage:** In addition to Hindi, Devanagari is used for Marathi, Tamil, Telugu, Malayalam, Kannada, and many other languages. It is important to consider the regional variation in the use of Devanagari to develop CAPTCHA that resonates with different linguistic communities.

**3.4 Security Concerns:** Devanagari Script-Based CAPTCHA Security Concerns Devanagari script-based CAPTCHA security concerns[18,19,20] are very important because CAPTCHA plays an important role in preventing automated attacks on online systems. The use of Devanagari scripts comes with its own set of challenges. However, addressing the security concerns is very important to make sure that the CAPTCHA is effective in preventing automated scripts and bots. Here are some of the security concerns related to Devanagari scripting-based CAPTCHA

**3.4.1 Vulnerability to OCR attacks[18]:** However, Devanagari characters are prone to OCR attacks if they are not distorted properly. A good CAPTCHA should use distortion techniques that make OCR software difficult to read and interpret.

**3.4.2 Machine learning-based attacks[18]:** Traditional CAPTCHA can be challenged by advances in machine learning. Devanagari Script-based CAPTCHA can also be challenged by adversaries' machine-learning techniques. As a result, there is a need for continuous innovation in the design of CAPTCHA

**3.5 Usability and Accessibility:** Devanagari script-based CAPTCHAs require usability and accessibility[16,18]. Usability refers to the ease with which users can interact with the CAPTCHA and complete it. Accessibility refers to the ability of people with disabilities to participate in the CAPTCHA. Let's take a closer look at the following: Usability and Accessibility of Devanagari Script-Based CAPTCHA[21]

**3.5.1 User-friendliness:** Take into account the fact that users are familiar with various Devanagari scripts and characters. Try to create a design that appeals to a wide range of users.

**3.5.2 Avoiding Unnecessary Complexity:** Make the CAPTCHA as easy as possible. Avoid making the CAPTCHA more complicated than it needs to be. This can confuse users or discourage them from signing up.

#### 4. PROPOSED SOLUTIONS

**Behavioral Analysis Challenges:** To add an extra layer of protection, incorporate challenges that include behavioral analysis, for example, based on user behavior.

**Continuous Monitoring:** Systematically track and analyze CAPTCHA performance. Update CAPTCHA designs regularly based on new threats and user feedback.

**Education and Awareness:** Educate users about CAPTCHA and how it impacts online security. Increase awareness of security best practices to improve user collaboration.

#### 5. CONCLUSION

Combining these suggestions, developers, and designers will be able to develop CAPTCHAs based on Devanagari scripts that are not only safe from automated attacks but also easy to use, culturally sensitive, and accessible to a wide range of users. With regular updates and ongoing engagement with the user community, these solutions will be able to address the ever-changing security challenges.

This paper provides insights to researchers, developers, and security professionals working with CAPTCHA systems with specific challenges related to the implementation of such mechanisms using the Devanagari script, as well as possible solutions and future direction for improving the efficiency and security of Devanagari-based CAPTCHA systems

#### References:

- [1] S. B. Patil, G. R. Sinha, and K. Thakur, Isolated Handwritten Devanagari Character Recognition using Four Descriptors and HMM. *International Journal of Pure and Applied Sciences and Technology, Volume 8, No.1*, 2012, 69-74.
- [2] R.Jayadevan, S. R.Kolhe, P. M. Patil and U. Pal, Offline Recognition of Devanagari Script: A Survey, *IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications and Reviews, Volume 41, No. 6*, 2011, 782-796.
- [3] Ahn L, Blum M and Langford J (2004) Telling Humans and Computers Apart Automatically. *Communications of the ACM*, 47(2):56-60.
- [4] Naor M (1998) Verification of a human in the loop or Identification via the Turing Test. <http://www.wisdom.weizmann.ac.il/~naor/PAPERS/human.pdf>.
- [5] Ahmedy, I.; Portmann, M.; Using Captchas to Mitigate the VoIP Spam Problem; *Second International Conference on Computer Research and Development*, 2010. pp: 136 – 140.
- [6] H. S. Baird and K. Papat. Human interactive proofs and document image analysis. *Proc. of 5th IAPR Int. Workshop on Document Analysis Systems (DAS 2002)*, vol. 2423 of LNCS, pp. 507–518, 2002.
- [7] Carnegie Mellon University, CAPTCHA: Telling Humans and Computers Apart Automatically. Available from: <http://www.captcha.net/> [Accessed: April 22, 2015].
- [8] Pope, C. & Kaur, K. (2005), "Is it human or computer? Defending e-commerce with Captchas", *IT Professional*, vol. 7, no. 2, pp. 43-49.

- 
- [9] H. S. Baird, M. A. Moll, and S.-Y. Wang. Scattertype: A legible but hard-to-segment CAPTCHA. *Proc. of 8 Int. Conf. on Document Analysis and Recognition (ICDAR 05)*, vol. 2, pp. 935–939, August–September 2005
- [10] H. S. Baird, A. L. Coates, and R. J. Fateman. Pessimprint: a reverse turing test. *International Journal on Document Analysis and Recognition (IJ DAR)*, 5(2–3):158–163, April 2003
- [11] M. Chew and H. S. Baird. Baffletext: A human interactive proof. *Proc. of SPIE-IS&T Electronic Imaging, Document Recognition and Retrieval X*, vol. 5010 of *Proceedings of SPIE*, pp. 305–316, January 2003.
- [12] J. Elson, J. R. Douceur, J. Howell, and J. Saul. ASIRRA: a CAPTCHA that exploits interest-aligned manual image categorization. *Proc. of 14th ACM Conf. on Computer and Communications Security (CCS 2007)*, pp. 366–374, October – November 2007.
- [13] Ritendra Datta, J. Li, and J. Z. Wang. IMAGINATION: a robust image-based CAPTCHA generation system. *Proc. of 13th ACM Int. Conf. on Multimedia (MULTIMEDIA 05)*, pp. 331–334, November 2005.
- [14] Kurt Alfred Kluever, Richard Zanibbi. Balancing usability and security in a video CAPTCHA. SOUPS '09: *Proceedings of the 5th Symposium on Usable Privacy and Security*. July 2009.
- [15] Yalamanchili, Sushma, and M. Kameswara Rao. "A framework for devanagari script-based captcha." *arXiv preprint arXiv:1109.0132* (2011).
- [16] Pate, Sanjay E., and R. J. Ramteke. "Design and Generation of Devanagari Script CAPTCHA: Imaginative Technique." *First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022)*. Atlantis Press, 2023.
- [17] Bodkhe, P. S., and P. E. Ajmire. "END-BAR DEVANAGARI CHARACTERS RECOGNITION USING SVM AND PNN CLASSIFIERS FOR CAPTCHA." *Vidyabharati International Interdisciplinary Research Journal* 13(1), ISSN 2319-4979, 257-264.
- [18] Kumar, Mohinder, and Sanjiv Kumar Jindal. "Devanagari CAPTCHA: For the Security in Web ." *Tuijin Jishu/Journal of Propulsion Technology* ISSN: 1001-4055, Vol. 44 No. 4 (2023)
- [19] Kumar, Mohinder, and Manish Kumar Jindal. "Benchmarks for designing a secure Devanagari CAPTCHA." *SN Computer Science* 2 (2021): 1-16.
- [20] Yalamanchili, Sushma, and Kameswara Rao. "DEVACAPTCHA-AFramework TO PREVENT BOT ATTACKS." *Acharya Nagarjuna University, Andhra Pradesh, India. sushma\_yalamanchili@ yahoo. co. in* (2011).
- [21] Kumar M, Jindal MK and Kumar M (2021a) A systematic survey on CAPTCHA recognition: types, creation and breaking techniques. *Archives of Computational Methods in Engineering*, Springer, 1–30.

## 7

**Stress Detection using Machine Learning Techniques****Mrs. M. M. Mohod<sup>1</sup>, Dr. P.M. Jawandiy<sup>2</sup>**<sup>1</sup>Post Graduate Department of Computer Science and Technology, DCPE, HVPM, Amravati, Maharashtra, India<sup>2</sup>Computer Science and Engineering, Pankaj Laddhad Institute of Technology and Management Studies, Buldhna, Maharashtra, India**Abstract**

Machine learning is used in healthcare for diagnosing the different diseases using various Machine Learning classification algorithms. Now a day, Stress becomes a common part of everyday life; it can be Positive or Negative. If it is negative many health issues occurs. These health problems associated with stress can prevent detecting mental stress earlier. The objective of this study is to identify the stress factors that affect the mental condition. Proposed techniques are applied on dataset with 2001 samples obtained from a Kaggel repository labeled Stress-Lysis. Parameters are human body humidity, body temperature and the number of steps taken by the user. Three different classifications labels of stress are performed, low stress, normal stress, and high stress. Support Vector Machine algorithm is applied. The performances of algorithms are evaluated on various measures like Accuracy, Precision, and Recall, F-measure.

**KEYWORDS:** Stress, Prediction, Support vector machine, Classification, Machine Learning,**I. INTRODUCTION**

As per Anxiety and Depression Association of America, stress become the most common illness in the U.S. 40 million adults are affected due to stress, every year and same case is happens in India. Now a day, Stress becomes common part of everyday life.

There are three types of stress. Short term stress is called acute stress which grow quickly still do not generally long lasting. Intense stress is called episodic acute stress which occurs during certain period of time. Very dangerous and harmful stress is called chronic stress. Detecting stress at earlier stage can prevent many health problems associated with stress. Many researchers are conducting experiments for diagnosing the different diseases using various classification algorithms of machine learning approaches like J48, SVM, Naive Bayes, Decision Tree, Decision Table etc.

Many researchers have proved that for diagnosing different diseases machine-learning algorithms works better. This research work focuses on people who are suffering from stress. In this work, Naive Bayes, Support Vector Machine, and Decision Tree machine learning classification algorithms are used. Experimental performance of all the three algorithms is compared on various measures and achieved good accuracy.

Proposed techniques are applied on dataset with 2001samples obtained from a Kaggel repository labeled Stress-Lysis. Parameters are human body humidity, body temperature and the number of steps taken by the user. Three different classifications labels of stress are performed, low stress, normal stress, and high stress.

The organization of this paper is as follows. In Section II Literature Review are presented. In Section III methodology carried out for this research are presented. In Section IV obtained results on the application of kaggle dataset are discussed. Finally in section V conclusions are drawn.

## II. LITERATURE REVIEW

Verma et al., 2020[1] proposed a model that predict stress level in students in technical education. The model uses Support Vector Machine and Logistic Regression algorithm based on linear kernel. The AUC-ROC accuracy of the Logistic model is 67% and SVM model is 86.84%.

Gedam and Paul, 2021[2] studied that various machine learning algorithms were applied to build classification models. The most common classifiers were Logistic regression, K-Nearest Neighbors, Random Forest, and Support Vector Machine. Mostly K-fold cross-validation is used for the classification models validation.

Ahuja and Banga, 2019[3] applied four classification algorithms Random Forest, Naive Bayes, Support Vector Machine, and K-Nearest Neighbour. Sensitivity, specificity, and accuracy are used as a performance parameter. The highest accuracy 85.71% was recorded by Support Vector Machine.

Pramod Bobade, 2020[5]proposed a model using WESAD dataset contains data from multiple physiological modalities like three-axis acceleration, respiration, electro dermal activity, electrocardiogram, body temperature, electromyogram and blood volume pulse. This model has achieved the accuracy of 84.32% on a three-class and 95.21% on a binary classification problem.

Elzeiny et al., 2019[7] implemented logistic regression, support vector machine and graph convolutional neural network. Models show promising cross validated classification accuracy with logistic regression 97.78%, support vector machine 93.67% and graph convolution neural network 89.58%.

Baheti & Kinariwala, 2019[8] proposed framework for detecting users psychological stress by using weekly social media data. For classification and prediction Supper Vector machine and Naive Bayes algorithm are used.

## III. METHODOLOGY AND EXPLANATION

### A. *Stress-Lysis Dataset*

For this research Jupyter Notebook used for performing the experiment. Jupyter Lab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality.

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

The main aim of this study is the prediction of the stress using the Jupyter Notebook by using the Stress-Lysis Dataset.

<b>Dataset</b>	<b>No. of Attributes</b>	<b>No. of Instances</b>
Stress-Lysis	4	2001

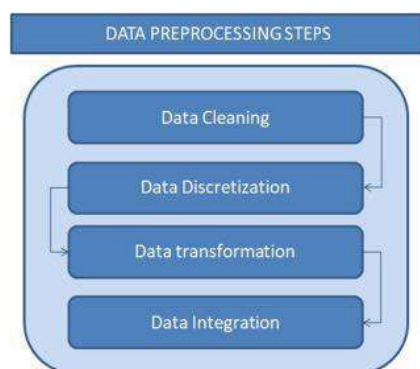


The proposed methodology is evaluated on Stress-Lysis Dataset, which is taken from Kaggle Repository. This dataset comprises of 2001 instances.

The dataset also comprises numeric 4 attributes Humidity, Temperature, Step count, Stress levels. Based on the human's physical activity, the stress levels of the human being are detected and analyzed. Three different classifications of stress are performed, low stress, normal stress, and high stress. Rachakonda et al., 2019[9]

### B. Preprocessing Data

Data preprocessing is a basic and primary step for converting raw data into useful information. In general raw data could be incomplete, redundant, or noisy. By data preprocessing, all these mentioned issues can be resolved and used for generating machine learning models.[1]



### C. Support Vector Machine

Support Vector Machine, abbreviated as SVM. SVM is a powerful supervised algorithm. It can be used for both regression and classification tasks, but generally, they work best in classification problems.

#### Multiclass Classification Using SVM

SVM doesn't support multiclass classification natively. It generally supports binary classification and separating data points into two classes. For multiclass classification, the same methodology used as binary classification problems. Multiclassification problem is breaking down into multiple binary classification problems.

The idea is to mapping data points to higher dimensional space to gain mutual linear separation between every two classes. This is called a *One-to-One* approach, which breaks down the multiclass problem into multiple binary classification problems. A binary classifier per each pair of classes.

Another approach one can use is *One-to-Rest*. This approach mainly splits the multiclass data as binary classification data so that the binary classification algorithms can be applied to convert binary classification data. According to the two breakdown approaches, to classify data points from  $n$  classes data set:

- In the *One-to-Rest* approach, the classifier can use  $n$  SVMs. Each SVM would predict membership in one of the  $n$  classes.
- In the *One-to-One* approach, the classifier can use  $n(n-1)/2$  SVMs.

## IV. RESULT ANALYSIS

After loading of the dataset, preprocessing is done to handle missing values and wrong values in the dataset to make dataset better. For prediction kernel based Linear Support Vector Machine algorithm is used. There are various kernel methods can be used with Support Vector Machine, but in this study linear kernel is used for the prediction of the stress. The dataset was splitted into training and test set. 80% of data is taken as training and remaining 20% is taken as test dataset. Support vector machine is giving an Accuracy 99 %, Precision 99 %, and Recall 100 %, F-measure 99%.

## V. CONCLUSION

One of the important real-world medical problems is the detection of stress at its early stage. In this paper, systematic efforts are made in designing a system which results in the prediction of stress. Support Vector Machine algorithm are applied and evaluated on various measures. Support vector machine is giving an accuracy of Accuracy 99 %, Precision 99 %, and Recall 100 %, F-measure 99%.

## References

- [1] G. Verma, S. Adhikari, V. Khanduri, S. Tandon, S. Rawat, and P. Singh, "Machine Learning Model for Prediction of Stress Levels in Students of Technical Education," in *Proceedings of International Conference on Applied Mathematics & Computational Sciences*, Aug. 2020, pp. 53–61, doi: 10.21467/proceedings.100.5.
- [2] S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/ACCESS.2021.3085502.
- [3] A. B. Ravinder Ahujaa, "Mental Stress Detection in University Students using Machine," *Procedia Comput. Sci.*, vol. 152, no. October, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.
- [4] P. Bobade and V. M., *Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data*. Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020) IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2 II., 2020.
- [5] P. Bobade and M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data," in *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*, Jul. 2020, pp. 51–57, doi: 10.1109/ICIRCA48905.2020.9183244.
- [6] S. Elzeiny, M. Qaraqe, Institute of Electrical and Electronics Engineers, P. Manjunath, S. S. Panicker, and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, 2019, doi: 10.1016/j.bbe.2019.01.004.
- [7] S. Elzeiny and D. M. Qaraqe, "A Machine Learning Approach for Detecting Mental Stress Based on Biomedical Signal Processing," Jun. 2019, doi: 10.5339/qfarc.2018.ictpd365.
- [8] R. R. Baheti and S. Kinariwala, "Detection and analysis of stress using machine learning techniques," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 335–342, Oct. 2019, doi: 10.35940/ijeat.F8573.109119.
- [9] L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel, "Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT."

## 8

## IoT Based- Soil Salinity Mapping and Smart Crop Recommendation System

Dr. Avinash B. Kadam<sup>1</sup>, Miss. Shubhangi D. Falke<sup>2</sup>,

Mr. Chandrakant R. Patorkar<sup>3</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

Corresponding author: E-mail: avinashkadam28@gmail.com<sup>1</sup>

<sup>2</sup>Research Scholar, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

E-mail: shubhangifalke96@gmail.com<sup>2</sup>

<sup>3</sup>Research Scholar, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

E-mail: chandrakantpatorkar7@gmail.com<sup>3</sup>

### Abstract:

With the increasing global demand for sustainable agriculture and the rise of the Internet of Things (IoT), precision agriculture has become a focal point for enhancing crop yield and resource utilization. This research paper introduces an innovative approach for addressing the challenges of soil salinity in agriculture through the integration of IoT technologies.

This research paper investigates the implementation of an IoT framework to address soil salinity challenges in agriculture. The proposed system integrates IoT technology for Soil Salinity Mapping and Smart Crop Recommendations. Through wireless sensor networks and advanced analytics, the system provides real-time soil salinity data and personalized crop recommendations, empowering farmers to make informed decisions for enhanced agricultural productivity in salinity-affected regions.

**Keywords:** IOT in Agriculture, Soil Sensor, Machine Learning, Smart Farming Technology, Soil Monitoring.

### Introduction:

The integration of Internet of Things (IoT) technology into agriculture has ushered in a new era of precision farming, offering farmers unprecedented capabilities to monitor and manage their fields in real-time. One critical aspect of agricultural productivity is soil health, with soil salinity being a major concern that can significantly impact crop yields.[1]This research focuses on the development of an IoT-based Soil Salinity Mapping and Smart Crop Recommendation System to address the pressing issue of soil salinity in agricultural landscapes.[2]The overarching goal is to provide farmers with actionable insights into their soil conditions and offer intelligent crop recommendations that are tailored to the specific salinity levels in their fields.[3]The research will involve deploying a network of IoT sensors across agricultural fields to continuously monitor soil salinity levels. These sensors will collect real-time data on various soil parameters, such as moisture content and electrical conductivity, enabling the creation of high-resolution soil salinity maps.[4]The integration of satellite imagery and remote sensing techniques will further enhance the accuracy and coverage of the mapping system.[5]In addition to soil salinity mapping, the research will focus on developing a Smart Crop Recommendation System.[6]The proposed IoT-based solution holds the potential to revolutionize traditional farming practices by providing farmers with actionable insights that empower them to make informed decisions. By mitigating the impact of soil salinity and optimizing crop choices, the research aims to contribute to sustainable agriculture, ensuring food security in the face of changing environmental conditions.[7]In the ever-evolving

landscape of agriculture, the integration of cutting-edge technologies has become pivotal for optimizing crop production and resource management. This research focuses on the innovative application of IoT in addressing the challenges posed by soil salinity in agriculture. By introducing an IoT-based Soil Salinity Mapping system coupled with a Smart Crop Recommendation system, we aim to empower farmers with real-time data and intelligent insights for making informed decisions, enhancing crop yields, and promoting sustainable farming practices.

### **Literature Review:**

In this paper the literature review section of a research paper on "IoT Based- Soil salinity mapping and Smart Crop Recommendation System" explore existing studies, theories, and findings related to IoT in agriculture, crop recommendation systems, and smart farming.

This paper describes the development of a complete IoT system for a smart farm with some main aims:1)Smart farming is a development that has emphasized information and communication technology used in machinery, equipment, and sensors in network-based hi-tech farm supervision cycles. Innovative technologies, (IoT) and cloud computing are anticipated to inspire growth and initiate the use of robots and artificial intelligence in farming.2)Soil salinity accumulates a high concentration of salts in soils that interfere with normal plant growth. Early detection and quantification of soil salinity are essential to effectively deal with soil salinity in agriculture. Soil salinity quantification and mapping at the irrigation scheme level are vital to evaluating saline soil's reclamation activity. Existing solutions of salinity mapping are costly, time-consuming, and inadequate for applications at the irrigation scheme level. (IoT) assisted salinity mapping at the irrigation scheme level is proposed to quantify and map the soil salinity in agriculture. The proposed IoT-assisted salinity mapping characterizes the soil salinity in terms of Electric Conductivity, pH, and Total Dissolved Salts. The accuracy of proposed IoT-assisted salinity mapping is evaluated against the standard method of salinity measurements. The proposed IoT-assisted salinity mapping is cost-effective, and portable, which is very useful for site-specific treatments and soil zones management in saline soils.3)This research will help learner and poor farmer as a guidance for cultivation of crops according to soil and climate condition by the use of modern technologies like Machine Learning (ML) and Internet of Things (IOT). The data regarding seeds and crops are collected with appropriate parameters like Soil types, Temperature, Moisture holding capacity of soil and Humidity which help to get prosperous growth. Moreover, to this we have developed a module by which farmer will be able to monitor the farm from a remote distance.4)Soil salinity accumulates a high concentration of salts in soils that interfere with normal plant growth. Early detection and quantification of soil salinity are essential to effectively deal with soil salinity in agriculture. Soil salinity quantification and mapping at the irrigation scheme level are vital to evaluating saline soil's reclamation activity. Existing solutions of salinity mapping are costly, and inadequate for applications at the irrigation scheme level. The proposed IoT-assisted salinity mapping characterizes the soil salinity in terms of Electric Conductivity, pH, and Total Dissolved Salts. The proposed IoT-assisted salinity mapping effectively observes impacts of reclamation activities in saline soil by frequent observation of soil salinity cost-effectively.5)Accurate monitoring of soil salinization plays a key role in the ecological security and sustainable agricultural development of arid regions.

### **Methodology:**

There are several steps that are taken and should provide a detailed and transparent description of the research design, data collection procedures, and analytical methods.

Research Design:

Define the overall research approach, whether it is experimental, observational, or a combination of both. Justify the chosen design in the context of the study's objectives.

**Data Collection:**

Describe the process of data collection. Outline the frequency and duration of data collection, specifying whether it was continuous or periodic. Provide information on how soil salinity data was recorded and transmitted from the sensors to the cloud platform also collect information about crops, soil, weather and other relevant factors from reliable sources.

**Smart Crop Recommendation System:**

Detail the components of the smart crop recommendation system. Explain how data related to environmental conditions, and other relevant factors were collected.

**Integration and Cloud Computing:**

Describe how the collected data, both from soil salinity mapping and crop-related factors, was integrated into a centralized cloud-based platform. Discuss the communication protocols employed for seamless data transmission and storage.

**Machine Learning Algorithms**

If machine learning algorithms were utilized for data analysis, provide an overview of the specific algorithms employed. Explain how these algorithms were trained and validated, and discuss any parameter tuning conducted.

### **Components:**

This research comprises several key components that work collaboratively to collect, process, and provide actionable insights to farmers. Here are the main components:

**IoT Sensors**

**Soil Salinity Sensors:** Measure the electrical conductivity of the soil to determine salinity levels.

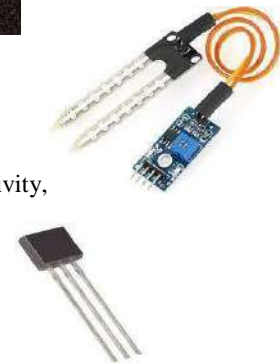
#### **i) Soil Moisture Sensors:**

It Measure the volumetric water content in the soil. It provides information about soil water availability, aiding in irrigation management and nutrient transport analysis.



#### **ii) Temperature Sensors:**

Measure the soil temperature. The importance of Soil temperature influences microbial activity, nutrient availability, and plant growth.



#### **iii) Soil Mapping Sensor:**

Soil mapping involves locating and identifying the different soils that collecting information about their location, nature, properties and potential use, and recording this information on maps and in supporting documents to show the spatial distribution of every soil.



**1] Data Communication Network**

i) **Wireless Connectivity:** Utilizes communication protocols (e.g., Wi-Fi, LoRa) for seamless data transmission between IoT sensors and the central data processing unit.

**2] Central Data Processing Unit:**

i) **Data Aggregation:** Gathers and aggregates data from the IoT sensors distributed across the agricultural field.

ii) **Data Processing:** Applies algorithms for real-time processing of sensor data to generate meaningful insights.

iii) **Data Storage:** Stores historical and real-time data in a secure and accessible manner, often leveraging cloud computing.

**3] Machine Learning Algorithms:**

i) **Crop Recommendation Algorithm:** Analyse soil salinity data, historical crop performance, and environmental factors to provide intelligent and personalized crop recommendations.

**4] Geospatial Technology:**

i) **Satellite Imagery:** Integrates satellite imagery and GPS data to enhance the accuracy and coverage of soil salinity mapping.

ii) **GIS (Geographic Information System):** Processes and visualizes geospatial data to create detailed soil salinity maps.

**5] User Interface (UI):**

i) **Dashboard:** Provides a user-friendly interface for farmers to access real-time data, soil salinity maps, and crop recommendations.

ii) **Customization Options:** Allows users to customize settings, view historical data, and receive alerts or notifications.

**6] Application Programming Interface (API)**

i) **Integration API:** Enables seamless integration with other agricultural management systems or external databases.

ii) **Communication with External Devices:** Facilitates communication with other IoT devices or machinery on the farm.

**7] Remote Monitoring and Control:**

i) **Remote Access Portal:** Allows farmers to remotely monitor field conditions, adjust irrigation, and receive insights from any location.

**8] Security Measures:**

i) **Encryption and Authentication:** Implements robust security protocols to protect data integrity and maintain confidentiality.

ii) **Access Control:** Manages user access and permissions to ensure data security.

**9] Educational Resources:**

i) **Knowledge Base:** Includes educational materials to help farmers understand the system, and make informed decisions.

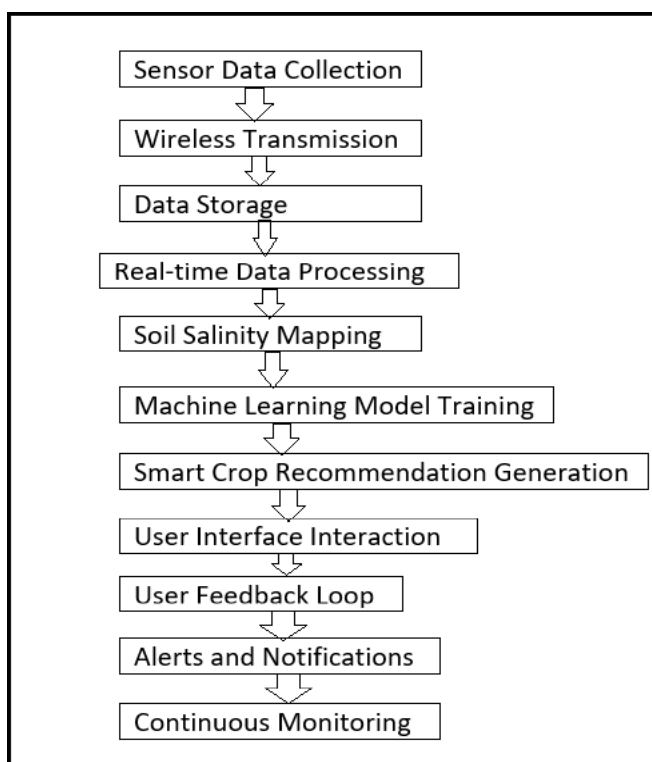
ii) **Training Modules:** Provides training resources to enhance farmers' skills in utilizing the technology effectively.

**10] Power Management:**

i) **Energy-Efficient Sensors:** Utilizes energy-efficient designs for IoT sensors, potentially incorporating renewable energy sources.

**ii) Battery Management:**

These components work in synergy to create a comprehensive IoT-based system that addresses soil salinity mapping and smart crop recommendations, providing farmers with valuable insights for efficient and sustainable agriculture.

**System Workflow:****Fig: System Workflow**

The system data flow in an IoT-Based Soil Salinity Mapping and Smart Crop Recommendation System involves the movement of information from various stages, starting from data collection to the generation of crop suggestion.

**1. Sensor Data Collection:**

Soil sensors collect data on parameters like soil salinity, moisture content, and temperature. Data is generated at regular intervals based on sensor readings.

**2. Wireless Transmission:**

Collected data is transmitted wirelessly from the soil sensors to a central server or cloud platform. Wireless communication modules facilitate the transmission of data.

**3. Data Storage:**

The transmitted data is securely stored in a centralized database or cloud storage.

**4. Real-time Data Processing:**

Advanced data processing algorithms analyze the real-time data received from the sensors. This processing step extracts meaningful insights and prepares the data for further analysis.

**5. Soil Salinity Mapping:**

Processed data is used to generate high-resolution soil salinity maps.

These maps visually represent the spatial distribution of soil salinity across the agricultural field.

**6. Machine Learning Model Training:** Historical data, including soil salinity maps and crop yields, is used to train machine learning models.

Models are trained to understand the relationships between soil conditions and crop performance.

**7. Smart Crop Recommendation Generation:**

The trained machine learning models, along with current soil conditions, generate personalized crop recommendations.

Recommendations consider optimal crop choices based on the analyzed data.

**8. User Interface Interaction:**

Farmers access the system through a user-friendly interface, such a mobile application.

The interface displays real-time soil salinity maps, historical trends, and crop recommendations to the user.

**9. User Feedback Loop:**

Users have the option to provide feedback on the system's recommendations through the interface.

User feedback is collected and considered for continuous improvement of the system.

## 10. Alerts and Notifications:

The system can send alerts and notifications to users based on critical changes in soil conditions or recommended actions.

Alerts may include suggestions for irrigation and changing the crop selection.

## 11. Continuous Monitoring:

The IoT system operates continuously, providing ongoing monitoring of soil conditions.

This work flow data ensures a seamless integration of information throughout the system, empowering farmers with real-time insights and intelligent crop recommendations.

### Application:

The research paper on "IoT-Based Soil Salinity Mapping and Smart Crop Recommendation System" presents a ground-breaking solution with diverse applications across the agricultural landscape.

### **Optimized Crop Selection:**

This application ensures optimal crop selection, leading to increased yields and reduced economic losses.

#### 1. Risk Mitigation and Crop Protection:

Timely detection of high soil salinity areas allows farmers to implement preventive measures, protecting crops from the adverse effects of salinity stress. This application minimizes crop losses and contributes to overall risk mitigation in agricultural operations.

#### 2. Data-Driven Decision-Making:

Farmers can analyze trends and patterns in soil salinity, enhancing their understanding of field conditions and supporting informed decision-making for future planting seasons.

#### 3. Remote Monitoring and Control:

The IoT framework facilitates remote monitoring and control, allowing farmers to access real-time data and receive alerts through mobile applications or web interfaces. This application enhances the convenience and efficiency of farm management, especially for farmers with geographically dispersed fields.

#### 4. Research and Development Collaboration:

This application contributes to a deeper understanding of soil salinity dynamics and supports the creation of improved crop varieties.

#### 5. Climate Resilience:

The system's ability to provide real-time information on soil conditions. Farmers can adapt their practices based on current conditions, enhancing their ability to withstand and recover from climate-related challenges.

### Result and Discussion:

In this section, a result is carried out for the difference and similarities between the proposed IoT-assisted salinity mapping and the standard method for salinity quantification.



**Fig: Sensor node prototype with sampling in the field**



**References:**

- 1) Bashir, R.N., Bajwa, I.S., Abbas, M.Z. et al.(IoT) assisted soil salinity mapping at irrigation schema level. *Appl Water Sci* 12, 105 (2022). <https://doi.org/10.1007/s13201-022-01619-1>.
- 2) Akhter, R.; Sofi, S.A. Precision agriculture using IoT data analytics and machine learning. *J. King Saud Univ.-Comput. Inf. Sci.* 2021, 34, 5602–5618.
- 3) Bashir, Rab & Sarwar, Imran & Abbas, Muhammad & Rehman, Amjad & Saba, Tanzila & Bahaj, Saeed & Kolivand, Hoshang. (2022). Internet of things (IoT) assisted soil salinity mapping at irrigation schema level. *Applied Water Science*. 12. 10.1007/s13201-022-01619-1. 4)Bhatnagar, Vaibhav & Chandra, Ramesh. (2020). IoT-Based Soil Health Monitoring and Recommendation System. 10.1007/978-981-15-0663-5\_1.
- 5) Roser, M.; Ritchie, H.; Ortiz-Ospina, E. World Population Growth. 2013. Available online: <https://ourworldindata.org/world-population-growth> (1 August 2022).
- 6) M. Pramanik, M. Khanna, M. Singh, D.K. Singh, S. Sudhishri, A. Bhatia, R. Ranjan Automation of soil moisture sensor-based basin irrigation system *Smart Agricult. Technol.*, 2 (2022), Article 100032, 10.1016/j.atech.2021.100032
- 7) A. A. Khan, M. Faheem, R. N. Bashir, C. Wechtaisong and M. Z. Abbas, "Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation," in *IEEE Access*, vol. 10, pp. 129505-129519, 2022, doi: 10.1109/ACCESS.2022.3228160.
- 8) Redmond R. Shamshiri, Siva K. Balasundram, Abdullah Kaviani Rad, Muhammad Sultan and Ibrahim A. Hameed Submitted: 06 February 2022 Reviewed: 23 February 2022 Published: 22 June 2022 DOI: 10.5772/intechopen.103898.

## Big Data Analytics In Health Care: A Reviewpaper

**Prof. Rana Afreen Sheikh**

**Prof. S. K. Totade**

Department of MCA, Vidya Bharati Mahavidyalaya, Amravati

### **ABSTRACT**

*The application of big data in health care is a fast-growing field, with many discoveries and methodologies published in the last five years. Big data refers to datasets that are not only big but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Moreover, medical data is one of the most growing data, as it is obtained from Electronic Health Records (EHRs) or patients themselves. Due to the rapid growth of such medical data, we need to provide suitable tools and techniques in order to handle and extract value and knowledge from these datasets to improve the quality of patient care and reduces health care costs. Furthermore, such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. This paper presents an overview of big data content, sources, technologies, tools, and challenges in health care. It also intends to identify the strategies to overcome the challenges.*

**KEYWORDS** *Big Data, Healthcare, Big data challenges, EHRs.*

### **1. INTRODUCTION**

Nowadays there is increasing in the details and data presented through the advancements in technologies and the internet. Anything ranging from consumer names and addresses to products available, to purchases made, to employees hired, etc. has become necessary for day-to-

daycontinuity. Withtheimprovementinstoragecapacitiesandtechniquesofdatacollection,enormous amounts of data have become easily available. Every second, more and more data is being produced and needs to be stored and analyzed in order to obtain value. Furthermore, data havebecome cheaper to store, so business companies and organizations need to get as much value as possible from the huge amounts of data collected daily.

Data sets increase rapidly because they are frequently gathered by many information-sensing devices such as mobile devices, aerial (remote sensing), software logs and records, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks [1]. Thus, big data is a field that explains methods to analyze, systematically obtain information from, and how to deal with data sets that are too large or complex to be dealt with by traditional data processing applications.

The health care industry is one of the most important industries. It is also one of the world's largest and fastest-growing industries it can produce and handles data at a staggering speed, but different electronic health records (EHRs) collect data in different structures: structured, unstructured, and semi structured. This variety can pose a challenge when seeking veracity or quality assurance of the data. The EHRs can provide a rich source of data, ready for analysis to improve our understanding of disease mechanisms, as well as better and personalized health care, but the data structures pose a problem to standard means of analysis. So, there is a need for converting the raw Data into significant and action able information by using big data analytics tools [2].

Big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software or popular tools and methods [3]. Accordingly, big data in healthcare is overwhelming not only because of its volume of data sets but also because of the variety of data types and the speed at which it must be managed.

The purpose of this systematic review is to provide a summarize of big data analytics in healthcare. First, we define and explain the definition of big data and the characteristics of big data analytics in the healthcare domain. Then we describe the big data types in healthcare. Third, we provide examples of big data analytics in healthcare. Fourth, we compile a list of challenges and opportunities faced by big data analytics in health care. Finally, we offer conclusions and future directions.

## 2. BACKGROUND

### 2.1. Defining Big Data

The concept of “big data” is not new, however, the way it is defined is continually changing. Many authors have provided big data definitions such as Zulkarnain et al. [4] define Big Data as “data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze”. Likewise, Kaislere et al. [5] say “Big data is data too big to be handled and analyzed bytr additional database protocols such as SQL”. Moreover, the authors in [6] present big data as a collection of data elements whose size, speed, type, and/or complexity require an attempt to use and discover new hardware and software tools to successfully store, examine, and visualize the data. Accordingly, Big Data points to large, complex datasets that are exceeding the capabilities of the traditional data management system to store , manage and process them.

### 2.2. Big Data Characteristic

As with all big things, if we want to manage them, we need to characterize them to organize our understanding. The three Vs (volume, velocity, and variety ) are known as the main characteristics of big data. These features are key to understanding how we can measure bigdata. The volume of the data refers to its size, and how huge it is. While the velocity points to the rate with which data is changing, or how often it is created. Finally, the variety involves several formats and types of data, as well as the different kinds of uses and ways of analyzing the data [7]. The characteristics are described below in Fig.1.

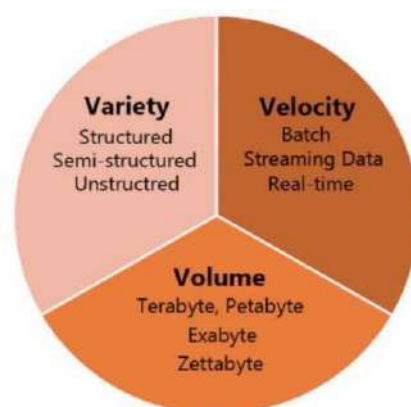


Figure1. The Big Data Characteristic

As shown in Figure.1. Big Data can be described by the following characteristic:

**Data volume:** This is the first and most important attribute of big data. Big data can be quantified by size in Tb soar PBs, as well as even the number of records, transactions, reports, or files. The volume of data used to play important role in storage and processing. However, many factors can contribute to the volume rise in data, it could amount to hundreds of terabytes or even petabytes of information generated anywhere. As displayed in [8], the number of data sources for an organization is growing day by day. And therefore, more data sources consisting of enormous datasets increase the volume of data, which needs to be analyzed. As noted in [8], Fig. 2 shows that the volume of data stored in the world would be more than 40 zetta bytes (  $10^{21}$  Byte) by 2020.

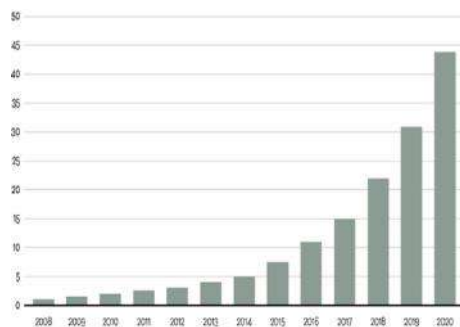


Figure2.Data volume growth by year in zetta bytes

- || **Data Velocity:** Points to the speed at which new data is generated and the speed at which data flows around, Hence, increasing speed in data processing, storage, and analysis by relational databases. Moreover, Velocity assists organizations understand their relative growth of their big data and how quickly that data reaches sourcing users, applications, and systems. Some activities are very important and require immediate responses, which is why quick processing maximizes effectiveness. For time-sensitive processes such as fraud detection, Big Data flows must be analyzed and used as they stream into the organizations to maximize the usefulness of the information. An illustration of data that is generated with great velocity would be Twitter messages or Facebook posts.
- || **Data Variety:** The next aspect of Big Data is its Variety. Which indicates the type of data that big data can contain. Big data is not always structured data. That means Big Data consists of any type of data, this data may be structured or unstructured such as text, sensor data, speech recordings, video, click streams, log files, and soon. Because Big Data contains data of different types and sources, Dealing with a variety of structured and unstructured data increases the complexity of both analyzing and storing Big Data. One of the goals big data is to employ technology to take this unstructured data and obtain an understanding of it.

### 2.3. Big Data in Health Care

In the healthcare field, the progress in information technology and the capability of storing more data have driven countries and governmental institutions to computerize health records and produced the Electronic Health Record (EHR) or Electronic Medical Record (EMR). Big data analytics in medicine and healthcare allows analysis of the large datasets from thousands

of patients, identifying clusters and correlation between datasets, Moreover improving predictive models using data mining techniques. As the health care industry focuses on improvements in order to save patients' lives, Big Data Analytics can play an important role in improving the services provided to health care by:

- Managing hospital performance
- Prevent epidemics, cure disease, and decrease costs.
- Increase transparency and efficiency in early disease diagnosis
- Enhancing clinical outcomes
- Engaging patients and family

### 3. RESULT

The literature included in this study contains essentially descriptive papers and studies. Based on the main research goals, the content from these studies was extracted and the papers were classified into many groups: Big Data analytics definition and concepts, sources of Big Data in healthcare, Big Data techniques for healthcare analytics, application and potential benefits of Big Data in healthcare and challenges in Big Data analytics in healthcare. The next section summarizes the conclusions in each of these categories.

#### 4.1. Big Data Analytics Concept

With the evolving of technology and the increasing numbers of data flowing in and out of organizations daily, there has become a demand for faster and more efficient ways of analyzing such data.

The author sin [9] explained that Big Data is in effective in a vacuum. So, its potential value is only obtained when used in decision making. To enable an organization to acquire knowledge and use it indecision-making, organizations need effective methods to apply large amounts of fast-moving data of various types and forms to analyze and benefit from it. The analytics concept refers to techniques used to analyze and acquire knowledge from big data. Thus, big data analytics can be viewed as a sub-process in the overall process of 'knowledge extraction' from big data.

As discussed in[10], Big data analytics refers to using advanced techniques and tools for analyzing and examining very large and various data sets that combine structured, semi-structured, unstructured data from various sources and in different sizes from terabytes to zetta bytes in aimsto obtain helpful information included within the data and will also help identify the data that is most important to the business and future business decisions. Instead of: hidden patterns, associations, market trends, and consumer preferences.

#### 4.2. Source of Health Care Big Data

Data that is obtained, collected, and stored in the healthcare industry may be are disorganized and distributed, coming from various sources and having different structures and forms. Health care Big Data involves data on physiological, behavioral, clinical, environmental illness, medical imaging, disease administration, medicine prescription records, nutrition, or exercise parameters [11]. However, most of the studies reviewed agreed on common sources of big data in the healthcare field, which are as follows:

- Electronic Health Records (EHRs):** An electronic copy of a patient's medical record thatismaintainedbytheserviceproviderovertime.TheEHRscanbecontainingdatarelatedt othe results of clinical and administrative meetings between the service provider (doctor,

□□□□□□□□ nurse, etc.) and the patient [12]. EHRs may include arrange of data including demo graphic Medical history, medication and allergies, immunization status, laboratory test results,

Radiology images, vital signs, personal statistics like age and weight, billing information, and Active medical problems [13].

**Electronic Medical Records (EMRs):** EMRs are similar to EHRs, they are digital records of patient health information; it is a digital version of a patient's information maintained in the formula chart, and it contains the patient's medical and treatment history from one clinic. Usually, this digital record stays in the doctor's office and does not get shared. If a patients witches doctors, his or her EMR is unlikely to follow. However, this paper chart is stored in clinician offices, clinics, and hospital data bases [14].

**Patient- Reported Outcomes (PROs):** Defined as a report coming directly from patients about their health condition and treatment which are based on a patient's perception of a disease and its treatment. This report includes arrange of outcomes such as symptoms, health status, and health-related quality-of-life [15].

**Data collected from wear able sensors:** The majority of wearable devices allow the collection of biochemical, physiological, and motion-sensing data such as (Heart rate, Steps walked, Blood pressure, etc.). So, it can collect patient health data and have data sharing capabilities [16]. The analysis of this type of data, when integrated with electronic health records, can support health monitoring and diagnosis for different chronic conditions.

**Data extraction from social networking tools (social media):** Patient posts on online social media such as Facebook, Instagram, Twitter, etc. can be extracted to obtain information about disease trends, patients' satisfaction, happiness, interests, and feelings. Twitter is a common example where data analytics methods have been used for disease monitoring and health-related trends (e.g. [17]).

#### 4.1. Big Data Analytical Techniques and Tools in Healthcare

Different types of healthcare data are difficult to analyze due to their dynamicity and complexity, such as medical images (X-ray, Magnetic Resonance Imaging MRI images), biomedical signals (EEG, ECG, EMG, etc.), audio records, multi-dimensional health care data, written prescriptions and structured data from EMRs and EHRs [18]. Moreover, there is a lack of analytical approaches that can handle such unstructured data and help decision making [19]. In this review, we summarize the literature that considers some of the analytical strategies and tools which can apply to health care and medicine.

As reported by (Asante-Korang and Jacobs, 2016) [20], there are 4 types of Big Data Analytics: Descriptive, Diagnostic, Predictive, and Prescriptive Analytics. According to the literature, predictive analytics is the most popular in the healthcare industry as they are used to detect early signs of patient deterioration, predict high-cost patients, re-admission, what might happen (when the patient's condition worsens), adverse events, and treatment improvement for diseases affecting the multi-organ system as discussed in [21, 22,23]. Moreover, Healthcare organizations have observed improved quality of care after adopting several Big Data analytics techniques that helped enhance the ability of the healthcare sectors to predict epidemics and treat disease. Table 1. Summarizes some of the Big Data Analytical Techniques used in healthcare.

**Table1. Summarizes some of the Big Data analytical techniques used in health care.**

#### 4.2. Big Data Analytics Challenges in Health Care

Big Data helps organizations, individuals, countries, and the world to create new growth opportunities, but it also poses significant challenges that could offset any potential gains, such as the loss of privacy and confidentiality, and the lack of appropriate IT infrastructure. Also, many of the big data tools are open source and free to use, which could provide the opportunity for intrusive operations, hackers, and data theft. Some literatures [32-38] discuss obstacles in the development of big data in health care applications. The key challenges are listed as follows:

1. **Privacy and Security:** Privacy and security are a key concern for individuals and corporations that hold information/data about people, products, activities, etc. Medical data obtained by healthcare providers from individuals and their medical records may contain private and confidential data [32]. Wherefore, protecting the patient's information must be handled with

enormous care from harm and hacker. When we use big data, many tools applied to analytics and data processes are open source and do not include all security measures [33]. Therefore, the primary justification for protecting personal privacy is to protect the interest of individuals. In order to overcome these challenges, some approaches are used to enhance the security level and obtain some confidentiality. First, Employing security measures, including strong encryption of data, validation of the source of data, access control, and authentication, where authentication is one of the measures for securing the data and maintaining confidentiality.

2. **Storage and Processing Issues:** Doubtlessly, the most obvious challenge associated with big data is simply storing and analyzing the huge amount of data. Nowadays, data grow significantly whenever a new storage technology is invented due to the huge amount of data collected and transferred by social media, healthcare providers, business transactions, and stock markets [34]. Moreover, this data is not just high on volume, but it also includes data of varied kinds that is generated periodically. With

Analytic Technique	Healthcare Application	Studies By
Cluster Analysis (CA)	-Identify cost change patterns of patients with end -stage renal disease(ESRD) who initiate dhemo dialysis (HD)by applying different clustering method.	ISMAIL etal., [21]
Data Mining	-Determination of epidemics; - Detection some diseases - Management of health care and measuring the effectiveness of certain treatments	Jothi etal., [22]
Graph Analytics	-Analysis of hospital performance across various quality measures	Nisar etal., [23]
Natural Language Processing (NLP)	-Extract clinical concept (e.g. diagnosis, procedure, and symptoms)from electronic medical record, patient discharge summaries, and lab report.	Gudivada etal., [24]
Neural Networks	-Prediction of patients future disease -Diagnosis of chronic diseases;	Wang etal., [25]
Machine Learning	-Microsoft's Inner Eye application employs machine learning to differentiate between tumors and healthy anatomy using 3D radiological images that assist medical experts in radio therapy and surgical planning, among other things.	Qiu etal., [26]

the rate of data explosion, the biggest challenge in dealing with is big data is that the present or traditional systems are unable to store and process data of this size and kind [35]. Therefore, the storage problem can be solved by making use of cloud computing. This would enable small and medium-sized hospitals and care organizations to eliminate cost and data storage issues.

3. **Data Ownership:** Data ownership represents a crucial and ongoing problem in big data applications in healthcare and other areas. Though peta bytes of medical records generally belong to the healthcare providers, governmental healthcare systems, or hospital in which they were created, but the information in it is not owned by them [36]. On the other hand, patients believe that they own the data. This dispute may be ended in the legal system to resolve the ownership issues unless healthcare providers receive written approval from patients before using data for experiences or research objectives.
4. **Skills Requirement:** A data analyst is a professional whose work involves collecting, cleaning, visualizing, and transforming or modeling raw data in to the blocks of information that are used by marketers, developers, and even healthcare providers[37]. One of the most important challenges in dealing with big data is the skills required for individuals to works in the big data field. A recent study [38] examined the required skills to deal with big data and concluded that the skills you need to work with big data will involve analytical capabilities.

## CONCLUSION

The paper first defined what is meant by big data. We presented various definitions of big data, highlighting the fact that size is only one dimension of big data. Other dimensions, such as velocity and variety, are equally important. The studies reviewed showed that big data in the health care industry is obtained from several sources such as results of medical examinations, hospital records, medical devices, and records of patients. For better treat disease and diagnosis in medical, the role of big data is one where it can construct better predictive models using tools with the ability to analyze and process this vast amount of data. Finally, a discussion has been made of some challenges that face individuals and organizations in the process of utilizing big data in healthcare, such as data ownership, privacy and security, storage and processing issues, and skills requirements.

## LIMITATIONS

While the proposed Review covers details about Big Data analytics and its applications in healthcare and medicine, however, we face a few limitations. First, the contents of this research consist of systematic review of the current state of Big Data technology in healthcare, but it does not get into consideration the technical details concerning the implementation and outcomes achieved in each of the studies reviewed. Second, there is heterogeneity in the documentation since

The literature includes various sources of information on the meaning of Big Data, methods of Big Data analytic, and their techniques and challenges in healthcare. Finally, despite the use of a Systematic strategy for review, the inclusion of studies on big data analytics in 'healthcare' for this Review was based on personal experience and knowledge, hence the cross-reference literature were also examined for this review.



## FUTURE OUTLOOK

Big data analytics in medicine and healthcare is a very encouraging process of integrating, examining, and analyzing enormous amounts of complex heterogeneous data with different types: biomedical data, medical data, electronic health records data (EHRs), and experimental data. The combination of such various data makes big data analytics weave many fields, such as bioinformatics, medical imaging, sensor informatics, medical informatics, health informatics, and computational biomedicine. As further work, we plan to study the various improvements in big data analytic systems and databases. Also, we will attempt to produce a new high-performance data management system by depending on open source platform such as Apache Hadoop Map Reduce, which can assist heterogeneous datasets and uses memory and other hardware resources in a more efficient way to reveal hidden patterns.

## REFERENCES

- [1] S. Singh, T. Firdaus, and A.K.Sharma,(2015)"SurveyonBigDataUsingDataMining",International Journalof Engineering DevelopmentandResearch, Vol.3, No.4, pp.135–143.
- [2] S. Dash, S. K. Shaky war, M. Sharma, and S. Kanshik, (2019) "Big data in healthcare: management, analysis and future prospects ",Journal of Big Data. Springer International Publishing, Vol. 6, No.1.
- [3] I. K. Subagja, N. Amaliyah, U. Hiermy, B. T. Rahardjo, E. L. Lydia, K. Shakar, and P. T. Nguyen,(2019) "Evaluation of big data analytics in medical science", International Journal of Engineering and Advanced Technology, Vol. 8, (6 Special Issue 3), pp.717–720.
- [4] N. Zulkarnain, M. Anshari, and A. Definition, (2016)"BigData :Concept,Applications,&Challenges",InternationalConferenceonInformationManagementand Technology,pp.307–31.
- [5] S. H. Kaisler, F. J. Armour, and A. J. Espinosa, "Introduction to the big data and analytics: concepts,(2016)techniques,methods,andapplicationsminitrack",ProceedingsoftheAnnualHawaiiInternationalConference on System Sciences, pp.1059–1060.
- [6] J. Amudhavel, V. Padmapriya, V. Gowri, K. Lakshmi priya, K. P. Kumar, and B. Thiagarajan, (2015)"Perspectives, motivations and implications of big data analytics. ",In Proceedings of the2015InternationalConferenceonAdvancedResearchinComputerScienceEngineering&Technology,pp.1-5.
- [7] M. Pospiech, and C. Felden, (2012) "Big data - A State-of-the-Art", 18th Americas Conference onInformationSystems2012, AMCIS 2012, Vol.5,pp. 3918–3928.
- [8] S. Brown, (2020) "Data Characteristics", A multidisciplinary journal of global macro trends BIG,Vol.3,No.6, pp. 19–27.
- [9] M. I. Razzak, M. Imran, and G. Xu, (2020) "Big data analytics for preventive medicine", Neural Computing and Applications. Springer London. Vol. 32,No.9, pp.4417-4451.
- [10] G.Chen, andM.Islam,(2019)"BigDataAnalyticsinHealthcare",Proceedings-20192ndInternationalConferenceonSafetyProduceInformatization,IICSPI2019,Vol.2015,pp.227–230.
- [11] S. G. Alonso, I. Torre-Diez, J. Rodrigues, S. Hamorioui, and M. L. Coronado, (2017) "A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector.", Journal of Medical Systems, Vol. 41, No.11,pp.183.
- [12] A. P. Ambinder, (2005) "Electronic Health Records By", Journal of Oncology Practice, Vol. 1, No. 2,pp. 57–63.
- [13] R. S. Evans, (2016) "Electronic Health Records: Then, Now, and in the Future", Yearbook of medical informatics ,pp. S48–S61.
- [14] Z.Liangetal.,(2014)"Deep learningforhealthcaredecisionmakingwithEMRs",Proceedings-2014IEEEInternationalConferenceonBioinformaticsandBiomedicine,IEEEBIBM2014,(Cm),pp.556–559.
- [15] M. Calvert et al., (2015) "Putting patient-reported outcomes on the Big Data Road Map", Journal of the Royal Society of Medicine, Vol.108, No.8, pp. 299–303.
- [16] M. Uddin, and S. Syed-abdul, (2020) "Data analytics and applications of the wearable sensors in health care: An overview", Sensors(Switzerland),Vol.20, No.5.
- [17] J. Zhang, N. Xue, and X. Huang, (2016) "A Secure System for Pervasive Social Network-Based Healthcare ",IEEE Access, Vol.4, pp.9239–9250.
- [18] A.Kankanhallietal.,(2016)"Bigdataandanalyticsinhealthcare:Introductiontothespecialsection",Information SystemsFrontiers, Vol.18, No.2,pp.233–235.

- 
- [19] B. Van Calster et al., (2019) "Predictive analytics in healthcare: how can we know it works?" *Journal of the American Medical Informatics Association*, Vol.26, No.12, pp. 1651–1654.
- [20] A. Asante-Korang, and J. P. Jacobs, (2016) "Big Data and paediatric cardiovascular disease in the era of transparency in healthcare", *Cardiology in the Young*, Vol.26, No.8, pp. 1597–1602.
- [21] A. ISMAIL, S. ABDLERAZEK, I. M. EL-HENAWY, (2020) "Big Data Analytics in Heart Diseases", *Journal of Theoretical and Applied Information Technology*, Vol.98, No.11, pp. 1970–1980.
- [22] N. Jothi., N. A. Rashid, and W. Husain, (2015) "Data Mining in Healthcare - A Review", *Procedia Computer Science*. Elsevier Masson SAS, Vol.72, pp.306–313.
- [23] M. U. Nisar, A. Fard, and J. A. Miller, (2013) "Techniques for graph analytics on big data", *Proceedings-2013 IEEE International Congress on Big Data*, Big Data 2013, pp.255–262.
- [24] V. N. Gudivada, D. Rao, and V. V. Raghavan, (2015) "Big Data Driven Natural Language Processing Research and Applications", *Handbook of Statistics*. Elsevier Inc, vol.33, pp.203-238.
- [25] Y. Wang et al., (2014) "Energy efficient neural networks for big data analytics", *Proceedings-Design, Automation and Test in Europe, DATE*, pp.2–3.
- [26] J. Qiu et al., (2016) "A survey of machine learning for big data processing", *Eurasip Journal on Advances in Signal Processing*. EURASIP Journal on Advances in Signal Processing, Vol.1.
- [27] M. R. Ghazi, and D. Gangodkar, (2015) "Hadoop, mapreduce and HDFS: A developers perspective", *Procedia Computer Science*. Elsevier Masson SAS, Vol.48, pp. 45–50.
- [28] A. B. Patel, M. Birla, and U. Nair, (2012) "Addressing big data problem using Hadoop and MapReduce", *3rd Nirma University International Conference on Engineering, NUI CONE 2012*, pp.6–8.
- [29] A. Alam, and J. Ahmed, (2014) "Hadoop architecture and its issues", *Proceedings-2014 International Conference on Computational Science and Computational Intelligence, CSCI 2014*, Vol.2, pp. 288–291.
- [30] K. Dwivedi and S. K. Dubey, (2014) "Analytical review on Hadoop Distributed file system", *Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit*, pp. 174–181.
- [31] I. A. T. Hashem et al., (2020) "MapReduce scheduling algorithms: a review", *Journal of Supercomputing*, Vol.76, No.7, pp.4915-4945.
- [32] A. L'heureux, K. Grolinger, J. F. Elyamany, and M. A. Capretz, (2017) "Machine learning with big data: Challenges and approaches", *IEEE Access*, Vol. 5, pp.7776-7797.
- [33] A. Ghazvini, and Z. Shukur, (2013) "Security Challenges and Success Factors of Electronic Healthcare System", *Procedia Technology*. Elsevier B.V., Vol.11, pp. 212–219.
- [34] M. Padgavankar, and S. Gupta, (2014) "Big data storage and challenges", *International Journal of Computer Science and Information Technologies*, Vol.5, No.2, pp.2218–2223.
- [35] J. Li et al., (2014) "The overview of big data storage and management", *Proceedings of 2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing, ICCI\*CC 2014*, pp. 510–513.
- [36] P. Kostkova, "Who Owns the Data?", *Open Data for Healthcare*, vol.4.
- [37] R. Nambiar et al., (2013) "A look at challenges and opportunities of Big Data analytics in healthcare", *Proceedings-2013 IEEE International Conference on Big Data, Big Data 2013*, pp.17–22.
- [38] A. Gardiner et al., (2018) "Skill Requirements in Big Data: A Content Analysis of Job Advertisements", *Journal of Computer Information Systems*. Taylor & Francis, Vol.58, No.4, pp.374–384.

## Study of Software Vulnerabilities Detection Tools and Techniques

**Mr. Y. V. Hushare<sup>1</sup>**

[thehushare.yv@gmail.com](mailto:thehushare.yv@gmail.com)  
Shri Shivaji Science College,  
Amravati (MS)

**Dr. Sudhir B. Jagtap<sup>2</sup>**

[sudhir.jagtap7@gmail.com](mailto:sudhir.jagtap7@gmail.com)  
Swami Vivekanand Mahavidhyalaya,  
Udgir (MS)

**Dr. U. S. Junghare<sup>3</sup>**

[usjunghare@gmail.com](mailto:usjunghare@gmail.com)  
Shri Shivaji Science College,  
Amravati (MS)

### Abstract:

Software Vulnerabilities are one of the crucial causes in computer security problems. It is one kind of flaw in a software system which can crash the system or give invalid output or can behave in an unintended way. Software Vulnerabilities are one of the major reasons for cyberattack. There are many Software Vulnerabilities detection techniques, some are machine learning and some are deep learning techniques. Some vulnerability techniques find vulnerability in source code rather than binary code and some techniques find vulnerability in binary code. Finding vulnerabilities in open-source software is a challenging task. In this paper focus is given on study of software vulnerability tools and techniques using machine learning and deep learning.

**Keywords:** Open-source software, vulnerability detection, machine learning

### 1. Introduction:

Software vulnerability has long been a severe but crucial research issue in cybersecurity [1–3]. These security vulnerabilities threaten the IT infrastructure of organizations and government sectors. There are increasingly more vulnerabilities being discovered. Multiple vulnerabilities released in the Common Vulnerabilities and Exposures were approximately 4,600 in 2010. However, it grows to approximately 153,955 in 2021. Software vulnerability [4–6], as a threat, is increasing in frequency, scale, and severity, which are similar to natural disasters; it may lead to unintended and severe consequences. Once vulnerability in a key system is exploited by attackers, millions of computer systems may be affected [7].

Software vulnerability can cause disastrous consequences for information security. Earlier detection of vulnerabilities minimizes these consequences. Manual detection of vulnerable code is very difficult and very costly in terms of time and budget. Therefore, developers must use automatic vulnerabilities prediction (AVP) tools to minimize costs. Recent works on AVP begin to use techniques of deep learning (DL).

The term open source refers to something people can modify and share because its design is publicly accessible. Open-source software is software with source code that anyone can inspect, modify, and enhance. Prime examples of open-source products are the Apache HTTP Server, the e-commerce platform of Commerce, internet browsers Mozilla Firefox and Chromium and the full office suite LibreOffice.

### 2. Vulnerabilities in open-source software's:

Open-source vulnerabilities are basically security risks in open-source software. These are weak or vulnerable code that allows attackers to conduct malicious attacks or perform unintended actions that are not authorized. In some cases, open-source vulnerabilities can lead to cyberattacks like denial of service (DoS). The WhiteSource Vulnerabilities Database listed the following top 10 vulnerabilities observed in Open-source software's [11].

Open-source software's	Vulnerabilities
Lodash	Prototype pollution security issue
FasterXMLjackson-databind	Mishandling of typing and serialization gadgets and their interactions
HtmlUnit	Code execution issues in Android-specific Rhino engine
Handlebars	Arbitrary code execution security threat
http-proxy	Vulnerable to DoS attacks
Decompress	A critical Arbitrary File Write vulnerability.
XStream	A remote code execution issue
Netty	Allowed malicious actors an opportunity to exploit the system
Spring Framework	File download attack it can give an attacker complete access and control of a victim's machine through a downloaded file
PyYAML	Untrusted YAML files are processed through the full_load method

### 3. Vulnerabilities Detection Tools:

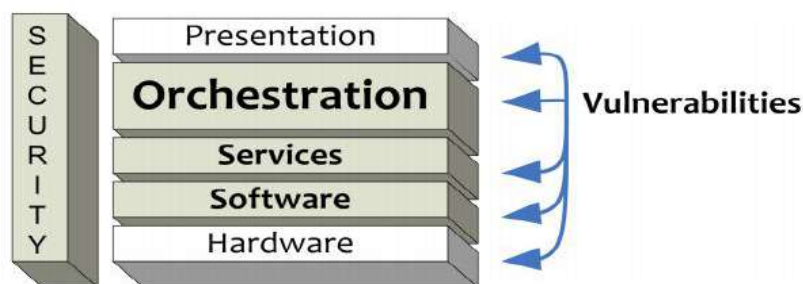
There are many vulnerability detection tools which are used to find vulnerabilities in Open-source software's. These vulnerability detection tools are as follows:

1. Static analysis tools: ITS4 is a tool based on lexical analysis technique, SPLINT (Secure Programming Lint), UNO, Checkstyle, ESC / Java (Extended Static Checker for Java), FindBugs, PMD is an open source, rule-based static detection tool.
2. Database tools: AppDetective by Application Security Inc
3. Developer Tools
4. Web Application Tools: AppScan DE by Watchfire, N-Stealth by N-Stalker iii. NTOSpider by NTObjectives iv. Spike Proxy by Immunity v. TestMaker by pushtotes vi. WebScarab by OWASP
5. Web Service tools: SOAPscope by Mindreef, SOA Test by Parasoft
6. Disassembler, Debugger, Decompiler tools
7. Binary/Bytecode Analyzer

### 4. Software Vulnerabilities detection techniques:

**4.1 Computer Emergency Response Team:** It is important to highlight the recent software vulnerabilities at regular basis to avoid the risks, but the disclosing the detected vulnerabilities is also area of concern from the developer's point of view. The market-based infomediary provides monetary rewards to identifiers for each vulnerability reported. When the market-based infomediary leaks vulnerability information, it is referred as the -based mechanism. When voluntary disclosure is low, encouraging a market-based mechanism with some regulation is a good idea. CERTtype infomediary and the market-based infomediary (where they exchange vulnerability information) that may lead to better results than the current environment, in which CERT and the market-based firm do not cooperate [7].

**4.2 Service Oriented Architecture:** An approach of extrapolating SOA vulnerabilities from previous vulnerability collections and classifying the resulting vulnerabilities according to their impact on the business process. A SOA vulnerability is a vulnerability present in such an environment. Vulnerabilities can exist in any of the SOA layers:



Focus lies on the orchestration layer, as it represents the biggest change a SOA brings. The orchestration layer contains the business process logic. Vulnerabilities which allow an attacker to modify either the activities or their sequence in a business process are our primary topic of interest. The typical approach to creating a SOA vulnerability classification would be to wait until a considerable amount of vulnerabilities has been collected, and then look for similarities based on which classes could be defined. Main approach is to examine existing vulnerabilities regarding their “survivability” in a SOA [8].

**4.3 Vulnerability Discovery Model:** The vulnerability discovery model (VDM), considered the impact of two factors that governed the rate of change of vulnerability discovered. The first factor deals with the number of installed base due to the rising popularity of a software and the second factor capture the decreasing phenomenon of the number of undetected vulnerabilities with time. First factor constitutes the vulnerabilities discovered with a detection rate  $r$  and the second factor represents additional vulnerabilities which are detected due to the influence of the vulnerabilities discovered by time  $t$ . The differential equation describing the discovery process can be modelled as

$$\frac{d\Omega(t)}{dt} = r(N - \Omega(t)) + s \frac{\Omega(t)}{N} (N - \Omega(t))$$

The predictability of the proposed model [9] has been assessed by fitting the VDM to an observed sample and evaluate the goodness-of-fit criterion of the fitted model on the observed samples to predict the future behaviour of the vulnerabilities. Non-linear least square methodology has been used to evaluate the estimation procedure on the security vulnerability data set. The results show a better insight about the vulnerability discovery process and revealed that it is better to use the proposed s-shaped model to estimate the vulnerabilities.

**4.4 Deep Learning-based Vulnerability Prediction:** The new features of Deep Learning (DL), especially in domains like computer vision and natural language processing, has sparked interest in using DL to detect security vulnerabilities automatically with high accuracy. Deep Learning-based Vulnerability Prediction (DLVP) techniques have been reported to detect security vulnerabilities with high accuracy; DLVP methods aim to detect unknown vulnerabilities in target software by learning different vulnerability patterns from a training dataset. Although Deep Learning is playing a vital role in vulnerability detection then also Key limitations of existing DLVP approaches that observed are: Data Duplication, Data Imbalance, Learning Irrelevant Features, Model Selection: Lack of Class Separation, etc. [10].

## 5. Conclusion:

In this paper overview of different open-source software’s and the various vulnerabilities found in their older versions are mentioned. Various tools which are used to detect open-source

software vulnerabilities were also discussed here. Most of the vulnerability detection techniques are machine learning and deep learning based. So the paper also provided concise study on open source software vulnerability detection techniques. In study it is found that different tools and techniques are used for different software's vulnerabilities. Each tool and technique have their pros and cons.

## 6. References:

1. YunfeiSu, Mengjun Li, Chaojing Tang, Rongjun Shen, An Overview of Software Vulnerability Detection, International Journal of Computer Science and Technology, Vol. 7, Issue3, July-Sept 2016.
2. Alfonso Fuggetta, Open source software- an evaluation, The Journal of Systems and Software 66 (2003) 77–90.
3. D Ming-Wei Wu, Ying-Dar Lin, Open-Source Software Development: An Overview, IEEE, Jun 2001.
4. Feller, Joseph; Fitzgerald, Brian, A framework analysis of the open-source software development paradigm, Proceedings of the twenty first international conference on Information systems, Brisbane, Queensland, Australia, 10-13 December., pp. 58-69.
5. Georg von Krogh, Eric von Hippel, Special issue on open-source software development, Research Policy 32 (2003) 1149–1157.
6. Sang-Yong Tom Lee, Hee-Woong Kim, Sumeet Gupta, Measuring open-source software success, Omega 37 (2009) 426 – 438.
7. Karthik Kannan, Rahul Telang, Market for Software Vulnerabilities? Think Again, Management Science, Vol. 51, May 2005, pp. 726-740.
8. Lutz Lowis and Rafael Accorsi, On a Classification Approach for SOA Vulnerabilities
9. Navneet Bhatt, Adarsh Anand, V. S. S. Yadavalli, Vijay Kumar, Modeling and Characterizing Software Vulnerabilities, International Journal of Mathematical, Engineering and Management Sciences, Vol. 2, No. 4, 288–299, 2017.
10. Saikat Chakraborty, Rahul Krishna, Yangruibo Ding, Baishakhi Ray, Deep Learning based Vulnerability Detection: Are We There Yet? IEEE Transactions on Software Engineering, VOL. TBD, 2020.
11. <https://soos.io/top-10-open-source-software-security-breaches>

## Trends in Recommendation Engine: A Systematic Review

Sachin N. Joshi<sup>1</sup>, Vivek A. Manwar<sup>2</sup> Dr. A. B. Manwar<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Sant Gadge Baba Amravati University.

[sachin.joshi106@gmail.com](mailto:sachin.joshi106@gmail.com),

<sup>2</sup>Research Scholar, Department of Computer Science, Sant Gadge Baba Amravati University.

[vivek.manwar007@gmail.com](mailto:vivek.manwar007@gmail.com)

<sup>3</sup>Associate Professor, Department of Computer Science, Sant Gadge Baba Amravati University.

### Abstract

We are living in information world, Information is emerging from various domains every day and from that scattered data metadata or user generated content or new algorithmic approaches are used for recommend information. Recommender systems or engines are used for filtering online information, which uses user's personal data, professional data, habits personalization trends, and user's activities on internet. Even though the recent recommender systems are giving specific recommendations, they suffer from various limitations and challenges. This paper highlights on various recent trends in the domain of recommender engine, concentrating on various applications like songs, social media, books, movies, products etc.

Keywords—Recommender system, Machine learning, Content-based filtering, Collaborative filtering, hybrid filtering, Deep learning.

### I. Introduction

Recommender systems are utilized in a various areas, and are most commonly recognized as playlist generators for video and music services like Spotify, YouTube and Netflix, product recommenders for services such as Flipkart Amazon, or content recommenders for social media platforms such as Facebook and Twitter. These systems can operate using a single or multiple inputs within and across platforms like books, news and search queries. There are also popular recommender systems for specific topics like restaurants and matrimony websites. Recommender systems have been developed to search research articles, collaborators, financial services, person you may know and life insurance.

### II. Literature review

**Anitha Anandhan et al.** [1] This paper aims to provide a complete review of the social media recommender system on research articles by using a methodological decision analysis in six aspects, including recommendation approaches, research domains, and data sets used in each domain, recommendation type, data mining techniques and the use of performance measures. Recommender system is developed based on user textual reviews, ratings, and comparative opinions. Recommender system for social media resources includes blogs, forums, social network websites, chatting sites, social bookmarking websites, video portals etc. Parameters considered in this paper are classification based on recommender approach, domain, dataset, data mining techniques, recommendation type and performance metrics. Matrix factorization technique mathematical model is used. **Arta Iftikhar et al.** [2] Collaborative filtering is the most commonly used technique for online recommendations. Collaborative filtering works by computing the interests of a user by gathering preferences or taste information of other users. Collaborative filtering is a personalization technique that counts on the set of ratings the user has provided for certain products and services. Issues discussed is accuracy of product recommendations using previous methods is not very promising. The randomness of this large amount of data, frustrate the users to choose what they

---

want. Proposed work aims to improve the accuracy of CF-based systems. Methodology Used for this work is triangle similarity method. An improved similarity measurement method based on the triangle similarity method to overcome the existing problems of similarity measure methods of collaborative filtering. The proposed method focuses on both common and non-common rating values of the user while producing recommendations. To improve accuracy, the proposed Item Recommender similarity measure is complemented with the rating preference behavior of users. Such systems employ social connection information and can help to mitigate more cold start and sparsity problem of Collaborative filtering systems.

**Peihai Zhao et al.** [3] With the rapid development of the Internet and the continuous progress of computer science and technology, Internet-based online trading has developed rapidly and improved the convenience of everyday life. Increasingly, people conduct business activities through online transactions and payment methods, and the development prospects for online transactions are quite broad. In addition, e-commerce trading systems have attracted considerable attention. Many studies have been conducted to analyze the behaviors of electronic commerce trading systems. Authors studied the existing user behavior analysis methods for e-commerce trading systems. Parameters considered are web browsing behavior, keystroke behavior, network transaction behavior, and mobile terminal behavior. Methodology used for this proposed system are pattern mining, process models. In this paper, authors provide an overview of advanced system behavior mining, consistency analysis and user behavior analysis for electronic commerce systems and discuss the differences and suitable application scenarios of these different methods. In future this research will help to work for personalized recommendation.

**Mohammed Alshammari, Olfa Nasraoui and Scott Sanders** [4] Recommender systems are being increasingly used on online platforms to predict the preferences of users and recommend relevant options to them. In particular, matrix factorization is a powerful recommendation model that can produce accurate recommendations, but unfortunately lacks transparency. This means that it fails to explain the reasons for its outputs, thus, it is called a black box recommender system. The aim of proposed work is to build a recommender system that can generate both accurate predictions and semantically rich explanations that specify the exact predictions. Issues discussed in this paper are black box and cold-start problems. For collecting data researcher conducted online and offline Survey from peoples. Matrix factorization mathematical model used for this work. The overall results indicate that the model succeeded in increasing the transparency of the system while keeping the error rate at a low level. Semantic knowledge graphs, linked data semantic distance matrix factorization and inferred fact style explanation are the methodologies used. The results clearly show that the proposed explanation style increased the user perception of system transparency, while being more effective in encouraging the user to accept the recommendation, leading to higher user satisfaction. In future predictions will be more accurate with the help of black box recommender system.

**Wenqian Liu et al.** [5] Online reviews and comments after product sales have become very important for making buying and selling decisions. Due to deceptive information fake reviews will affect such decisions, leading to financial losses for the consumers. Authors had analyzed the characteristics of review data using a crawled Amazon China dataset. Fake reviews based on temporal features of reviews and comments over E-commerce websites is the issues discussed by authors. Spamming detection and the detection of fake reviews using text mining techniques are the previous work done by another researcher. Parameters considered for work are number of reviews are total number of time slots. Mathematical model used is measurement metrics. Researchers apply three baseline methods, i.e., ARIMA, LOF and SVM to detecting outlier products using the crawled dataset. As time of interval increases the Accuracy rate is decreases.



Methodology used are temporal feature extraction, isolation forest algorithm and baseline methods. In this paper, authors studied the review records of online shopping sites and proposed a novel approach for the detection of fake reviews of products. By analyzing the temporal trends of reviews and comments Review outlier detection method detects the outlier products. This paper did not indicate clearly when a product has the highest probability of being involved in fake reviews and comments, which is another interesting piece of future work.

**Anusha Prabakaran, Min Chen** [6] With near global access to the Internet and growing popularity of online shopping, there is a rapid increase in the percentage of shopping through online stores. Unlike the traditional retail models, purchase decisions at these online stores are often based on product reviews, which should be made by real consumers who reveal their honest experiences of a product. However, due to the drive for profit or winning over competitors, some spammers attempt to bias product reviews by writing spam comments to intentionally deceive consumers. Unfortunately, spammers may write false reviews i.e. spam reviews growing or deprecating a product, which can mislead potential customers and negatively affect revenues of many genuine organizations. By running the test on a MacBook Pro, the system is able to generate the credibility report for all the product categories near about more than 2 million reviews within an hour. This makes the system usable for online website deployment. S-H-ESD algorithm is used as methodology. Credibility Score is the parameter used while proof. In this paper, authors present a multi-dimensional analysis to detect the spam reviews. It generates credibility reports for given products detection based on three methods: detection of duplicate reviews, detection of anomaly in review count and rating distribution, and detection of incentivized reviews. All the methods provide useful information, serving as an overlay to enable the discovery of fake reviews. In future for better recommendation this type work is important.

**Pei Yin and Liang Zhang** [7] With the rapid development of network technology, the number of digital images is growing at an alarming rate, people demands for information gradually shift from text into images. It is tough for users to quickly find the images they are interested in from the large number of image libraries. This recommend the more related personal information and image data in the image database to users, so that users can get the most interested picture data, that is, recommend the closest to the interesting picture to users. Issue identify in this work is when we search any image on web at that time it may show the non-relevant images, to solve this problem image recommendation is needed. Previous work done is Novel up-down cultural convolutional neural network (co-cnn), image retrieval. Parameters considered for this work are cross-layer context, global image-level context, semantic edge context, super-pixel in-context and cross-super-pixel neighborhood context. Methodology used are image retrieval algorithm, implicit support vector machine. Authors proposed image retrieval model, compared with the traditional content-based image retrieval technology, using improved local sensitive hash algorithm, to deep learning approach to study hash function to generate a binary hash code, binary hash code under the hamming space similarity measure to speed up the speed of image retrieval, in theory, should have good retrieval effect. This paper proposes and implements an image analysis recommendation algorithm based on implicit support vector machine. Different from existing image recommendation methods, this personalized image recommendation method combines image text information and image content information. Proposed image recommendation method based on implicit support vector machine can meet the personalized needs of different users.

**Marwa Hussien Mohamed, M. Helmy Khafagy and M. Hasan Ibrahim** [8]

Recommender systems are playing an important role to help e-commerce growing in many applications at WWW. Also, a recommender system is one of the most important research areas today's because it helps users to find their interest in the internet. Recommender systems used to predict the rating that the user will give to an item if this the first time to see on the site by

using information filtering systems from the user's rating list history on the site or by finding the item's similarity specification or by finding the common interests using demographic information from their profile. Issues discussed are cold start problem, sparsity problem, scalability, over specialization problem, diversity, novelty, serendipity, privacy, shilling attacks, gray sheep. Parameters considered are content analyzer, profile learner, filtering component. Some methodologies for mining and recommendation are outlier detection, classification, association analysis, regression analysis. Here, recommender engine software is required as tool. Recommender system used to recommend items to the user according to their interests and previous items rate list. In this paper, researchers discuss four techniques in recommender systems and list advantages and disadvantages for everyone. In future, what about using Big data algorithm like map reduce to increase performance in recommender systems.

**Maria N. Moreno et al.** [9] Web recommender systems are used in many application domains to predict consumer preferences and assist web users in the search for products or services. Recommender machines implemented in current systems have different levels of complexity, ranging from those that recommend products based on associations between them in previous transactions, those that sort recommendations based on evaluations that users provide about products and similarity between user preferences. The latter, known as collaborative filtering methods, are the most successful; however, some important drawbacks in them have been reported, especially in traditional approaches based on nearest neighbor algorithms, which show serious performance and scalability problems. In addition, the great number of evaluations needed by these methods in order to provide precise recommendations causes the sparsity problem when evaluations from users are insufficient. From the traditional collaborative filtering approaches to sophisticated web mining techniques numerous methods have been proposed to provide users with more and more effective recommendations, , however some important drawbacks are still present in current recommender systems. Issues discussed in this paper are scalability, sparsity, first rater and cold start problems. Previous work done are content based recommendations, collaborative recommendations, demographic recommendations and hybrid approaches. Parameters considered are content analyzer, classifiers. Some methodologies for Mining and recommendation are Outlier detection, Classification, Association analysis, Regression analysis. Movie Lens database is used for testing the new framework designed. In this work a recommendation framework specially talked to overcome jointly these weaknesses are proposed. The proposal contains combining web mining methods and domain specific ontologies in order to induce models at two abstraction levels. From data without semantic information the lowest level models are built. A case study with data from Movie Lens database was used to validate the framework. In this paper researchers do the case study and design movie recommendation framework in which they solve the cold start problem for recommender systems.

#### **Issues in Recommendation system:**

1. Cold Start Problem: -This problem occurs when new users enter the system or new items are added to the catalogue.
2. Synonymy: -Synonymy arises when an item is represented with two or more different names or entries having similar meanings.
3. Shilling Attacks: - Giving false ratings on some items either to increase the item popularity or to decrease its popularity. What happens if a malicious user or competitor enters into a system and starts? giving false ratings on some items either to increase the item popularity or to decrease its popularity.
4. Privacy: -Feeding personal information to the recommender systems results in better recommendation services but may lead to issues of data privacy and security. Users are reluctant to feed data into recommender systems that suffer from data privacy issues.

5. Limited Content Analysis and Over specialization: -Content-based recommenders rely on content about items and use recommendation system to be processed by information retrieval techniques. The limited availability of content leads to problems including over specialization. Some more issues are Grey Sheep, Sparsity, Scalability, Latency Problem, Evaluation and the Availability of Online Datasets, Context Awareness etc.

### III. Conclusion

It has been observed from the literature reviewed that recommendation systems are designed which uses different approaches which includes K-Means, C-Means, k-Nearest Neighbor, decision tree, clustering, regression, heuristic approaches, neural networks and association rule mining. With respect to the issues discussed in this domain, there is a better scope to apply proper recommendation framework with appropriate web content mining approach in information system which will eventually be an efficient way of searching desired data and will be applicable in business, institutions, governments, organizations and many sectors related to information world for decision making and developing business intelligence.

### IV. References

- [1] Anitha Anandhan et al., "Social Media Recommender Systems: Review and Open Research Issues", Volume 6, IEEE 2018
- [2] Arta Iftikhar et al., "An Improved Product Recommendation Method for Collaborative Filtering", Volume 8, IEEE 2020
- [3] Peihai Zhao et al., "Behavior Analysis for Electronic Commerce Trading Systems: A Survey", Volume 7, IEEE 2019
- [4] Mohammed Alshammari, Olfa Nasraoui and Scott Sanders, "Mining Semantic Knowledge Graphs to Add Explainability to Black Box Recommender Systems", Volume 7, IEEE 2019
- [5] Wenqian Liu et al., "A Method for the Detection of Fake Reviews Based on Temporal Features of Reviews and Comments", IEEE 2019
- [6] Anusha Prabakaran, Min Chen., "Product Review Credibility Analysis", Computing, Networking and Communication, IEEE 2019
- [7] Pei Yin and Liang Zhang, "Image Recommendation Algorithm Based on Deep Learning", Volume 8, Special Section On Gigapixel Panoramic Video with Virtual Reality, IEEE 2020
- [8] Marwa Hussien Mohamed, M. Helmy Khafagy and M. Hasan Ibrahim, "Recommender Systems Challenges and Solutions Survey", IEEE 2019
- [9] Maria N. Moreno et al., "Web mining based framework for solving usual problems in recommender systems- A case study for movies recommendation", Neurocomputing 176 72-80

## 12

**Machine Learning Tools for Data Science: A Review****Sagar A. Durge**

Research Scholar, Department of  
Electronics and Computer Science,  
PGTD, RTM, Nagpur University,  
Nagpur, M.S, India  
[sagardurge40@gmail.com](mailto:sagardurge40@gmail.com)

**Dr. Kishor M. Dhole**

Assistant Professor, Department of  
Computer Science,  
Seth. Kesarimal Porwal College of  
Arts and Science and Commerce,  
Kamptee. Nagpur, M.S, India  
[km.phd108@gmail.com](mailto:km.phd108@gmail.com)

**Abstract:**

Artificial Intelligence(AI), Machine Learning(ML), and Deep Learning(DL) are the buzzwords that have been grasping the interest of many researchers. As data usage gradually increases day by day. Data has been processed, organized, structured, and presented systematically. Data scientists need to manage and handle data and use them efficiently hence there is more demand for ML tools that work faster and better. This paper introduces various machine-learning tools that can be used at different phases of development to build a strong Machine-Learning model. The tools are developed using various methods, and various software libraries, programming languages, and algorithms. The tools can have a variety of different features. So, it is difficult to choose a tool for doing the application. A total of 14 ML tools were investigated to facilitate the selection of available tools. The general characteristics of the examined tools are described. Similar and different characteristics of the tools have been seen. It is aimed at making choices within the existing ML tools to help researchers.

*Keywords: Machine Learning, Data Science, Data Analytics, Machine Learning tools, Machine Learning Applications*

**I. Introduction**

Nowadays data is crucial and analysis and processing of data are much more important for Data Scientists to manage and handle data and use them efficiently hence there is more demand for Machine Learning (ML) tools that work faster and better. ML offers a wide range of tools, techniques, and frameworks to address the current challenges. As data usage gradually increases day by day. ML techniques and their applications are in use in day-to-day activities, such as health care, marketing, finance, logistics, banking, agriculture, and so on. ML models have become sophisticated and complex as a large amount of data is involved to solve problems as well as searching, advertisements, and YouTube, a multi-disciplinary field that has become synonymous with technological advancements and data handling challenges[1][2]. ML is the fastest-rising arena in computer science. ML techniques and tools have become ubiquitous in the process of analyzing data for diverse purposes. The underlying ML models are used, for instance, to predict future events based on the data at hand ML aims to develop algorithms that can learn and progress over time and can be used for predictions. Machine Learning practices are widely used in various fields. Machine learning (ML) and Data Analytics (DA) are proposed techniques that can help extract information and find valuable patterns within the collected data. Machine learning (ML) uses learning algorithms to learn from the data available.

ML uses data mining techniques to extract information from huge-size datasets. ML and Data Mining techniques explore data from end to end to find the hidden patterns inside the dataset [5]. Machine Learning and data mining algorithms have been deployed in various fields such as Computer networking, the travel and

tourism industry, finance, forecasting, the telephone communication industry electric load forecasting, and so on. [3] [4].

## II. Overview of Machine Learning Tools

**TABLE I**  
**Tools used for Model Development in Python**

S N	Tool Name	Developer & Year	Platform	Description	Features
1	Numpy	Travis Oliphant (2005)	Cross-Platform	Numpy very essential package for scientific computing. It comes with an N-dimensional array of homogeneous data types. It has a fixed size at the time of creation and elements are required to be the same type which makes this efficient with scientific computing. Supports Fourier transforms, mathematical functions, linear algebra methods, and random number generators. Numpy contains pre-compiled C code which is optimized. [5]	<ul style="list-style-type: none"> <li>i) High performance</li> <li>ii) Integration code from C/C++ languages.</li> <li>iii) Multidimensional container.</li> <li>iv) Broadcasting function.</li> <li>v) Work with varied databases.</li> <li>vi) Additional linear algebra functions.</li> </ul>
2	SciPy	Travis Oliphant, Pearu Peterson, Eric Jones (2001)	Cross-Platform	SciPy used for mathematical, scientific, and technical computing. SciPy ecosystem tools include a variety of data management and computation tools as well as efficient research and high-performance computation. Scipy library currently has a BSD license. It has high-level python commands to manipulate and visualize data. [5]	<ul style="list-style-type: none"> <li>i) It connects to the database quickly and flawlessly.</li> <li>ii) The SciPy library offers modules for linear algebra, image optimization, integration interpolation, special functions, Fast Fourier transform, signal and image processing, Ordinary Differential Equation (ODE) solving, and other computational tasks in science and analytics.</li> </ul>
3	Scikit-Learn	David Cournapeau (2007)	Cross-Platform	SkLearn is Python numeric and scientific library built to work with NumPy and Scipy. It makes use of Numpy and is widely used in high-speed linear algebra and operations. It has a significant method for analyzing predictive data. [5]	<ul style="list-style-type: none"> <li>i) It is the most popular Python ML library for developing ML algorithms. It has a wide range of supervised and unsupervised learning algorithms. The library can also be used for data mining and data analysis.</li> </ul>

					<p>ii)It has the following features:</p> <ul style="list-style-type: none"> <li>a) Clustering</li> <li>b) Regression</li> <li>c) Classification</li> <li>d) Dimensionality reduction</li> </ul> <p>iii)It integrates Matplotlib, Numpy, Scipy etc.</p>
4	Tensor Flow	Google Brain Team (2015)	Cross-Platform	<p>TensorFlow a popular framework for developing ML applications. It has a flexible architecture with which it can run on a variety of computational platforms CPUs, GPUs, and TPUs. It constructs models using different data flow graphs and also enables the developers to build large-scale neural Networks with several layers. [6]</p>	<ul style="list-style-type: none"> <li>i) Can create large-scale deep learning models with many layers.</li> <li>ii) Comprehensive control on developing a machine learning model and robust neural network.</li> <li>iii) It can run multiple CPUs and GPUs.</li> <li>iv) Supports abundant extensions and libraries for solving complex problems.</li> </ul>
5	Keras	Francois Chollet (2015)	Cross-Platform	<p>Keras is a python deep-learning library. It provides an interface to artificial neural networks like the TensorFlow library. It is the most popular deep learning framework. It is the best framework for quickly experimenting with deep neural networks. [6]</p>	<ul style="list-style-type: none"> <li>i) It uses a global state to implement functional model building API and uniquely auto-generated layer names.</li> <li>ii)It distributes the computing power of training a deep learning model on cluster of TPU &amp; GPU.</li> <li>iii)It simplifies the coding necessary for deep neural network.</li> <li>iv) It has plenty of tools to work on image and text data.</li> </ul>
6	Pandas	Wes McKinney (2008)	Cross-Platform	<p>Pandas is an open-source foundation Python software library for data analysis with support for fast, flexible, expressive data structures. It is used for manipulating time series and numeric tables. [3]</p>	<ul style="list-style-type: none"> <li>i) Supports different file formats like CSV, MS-Excel, SQL, and Fast HDF5 formats.</li> <li>ii) Flexible in reshaping and pivoting datasets.</li> <li>iii) Fancy indexing mechanism and integrated indexing mechanism.</li> <li>iv) Intelligent label-based slicing.</li> </ul>

					v)Automatic label alignment and manipulation of messy data.
7	PyTorch	Facebook AI Research Group(2016)	Cross-Platform	PyTorch is an optimized version of torch library which is used for deep learning. It uses CPUs and GPUs. The main motto of PyTorch is Computer Vision and Natural Language Processing. The interface of PyTorch is more enhanced and dynamic with Python. [5]	i) Easy, efficient, and interoperability to leverage the rich ecosystem of Python libraries as a part of a user program. ii) Writing programs in PyTorch is more Pythonic, in turn, it will be easy for Data Scientist. iii)Type-based automation differentiation systems developed to build a deep neural network.
8	OpenCV	Gray Bradsky (2002)	Cross-Platform	OpenCV stands for open-source computer vision, which is designed to work out the problem of computer vision. The main motto is to provide real-time computer vision. It is also used for video capture and analysis. It is written in C++. It supports Python, Java, and other programming languages. [6]	i)General Object Recognition. ii)Edge detection. iii) Feature matching for object recognition . iv) Color filtering. v)Subtracting the background from images. vi) Gesture recognition. vii) Motion tracking.
9	NLTK	Steven Bird and Edward Loper (2001)	Cross-Platform	NLTK is used to build Python programs that work with data of language used by humans. This platform also reinforces research and NLP for human language composed in Python. It has a user manual kind of book, that explains the fundamental concepts behind the language processing task of toolkit. [5]	i) NLTK is a community-driven project. So it has huge community support. ii) It also contributes the following a) classification b) tokenization c) stemming d) tagging e) parsing f) Semantic reasoning functionalities.
10	Google Auto ML	Google (2018)	Cloud	Google Auto ML is a cloud platform that is used to design ML models in Google Cloud and then we can integrate the model in our application. We can create a model using GUI. By using it we can create the ML models.[6]	i)AutoML Natural Language (this is used for text and documents). ii) AutoML Tables. iii) AutoML Translation iv)AutoML Video Intelligence. v) AutoML Vision and much more.

**TABLE II**  
**Tools used for Data Visualization in Python**

S N	Tool Name	Developer	Platform	Description	Features
1	Matplotlib	John D. Hunter (2002)	Cross-Platform	Matplotlib is a Python package that is used to depict data visually. It contains different methods which are used to represent data in different graphical forms. By representing the data graphically, we can gain more knowledge about data than seeing the data as it is. It contains an OO API which is used to process the embedding of the plots into the apps that are developed by the GUI toolkit. [5]	By using Matplotlib we can draw the following: i) Histogram ii) Scatter plot iii) Line plot iv) Image plot V) 3D plot vi) Pie chart and many more.
2	Seaborn	Michel Waskom (2017)	Cross-Platform	Seaborn is a Python library that is used to develop the statistical graph from the data. It is developed on the top of Matplotlib. It has many options to customize the plot which is drawn by using Matplotlib. Seaborn has many default options for attractive plots. Seaborn is like a statistical plotting frontend for Matplotlib.[4]	i) Updates plots ii) KDEplots iii) Pairplot iv) Violin Plot v) distplot vii)swarm plot
3	Tableau	Pat Hanrahan, Christian Chabot, and Chris Stolte (2003)	Cross-Platform	Tableau is a data visualization software that mainly focuses on business intelligence. Tableau won CODie award for 'Best Business Intelligence' in 2008. It has different software products like Tableau Desktop, prep builder, public mobile server, Vizable, and reader CRM. [5] [4]	i)Tableau dashboard which is used to combine different reports into one. ii)Collaboration and sharing. iii) Cloud storage support. iv) Third-party integration iv)Support all the advanced visualization techniques. v) Drag and drop interface
4	Plotly	Alex Johnson (2019)	Cross-Platform	Plotly Python package is an open-source, interactive plotting framework that supports many different charts for a variety of statistical, scientific, geographical, finance, and 3D applications. It offers enterprise products	By using Plotly we can draw the following: i) Scatter plot ii) Dot plot iii) Filled area plot iv) Box plot v) Tree plot



				like Dash Enterprise, Chart Studio Cloud, Chart Studio Enterprise, and Data Visualization libraries. [5] [2] [7]	
--	--	--	--	--	--

### III. Conclusion

This paper discussed different types of tools used in Machine Learning for various purposes like training, modeling, and visualization of data and also focused on different techniques to solve different problems. There are various tools for different tasks—Matplotlib for data visualization, SciPy library for scientific communication, and so on. Nowadays data is crucial and analysis and processing of data is much more important so we require some tools to do such tasks. So we require tools for that purpose. As data usage gradually increases day by day. Data scientists need to manage and handle data and use them efficiently hence there is more demand for Machine Learning tools that work faster and better.

### References

- [1] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, 'AN OVERVIEW OF MACHINE LEARNING', in *Machine Learning*, Elsevier, 1983, pp. 3–23. doi: 10.1016/B978-0-08-051054-5.50005-4.
- [2] B. Nithya and V. Ilango, 'Predictive analytics in health care using machine learning tools and techniques', in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai: IEEE, Jun. 2017, pp. 492–499. doi: 10.1109/ICCONS.2017.8250771.
- [3] S. Angra and S. Ahuja, 'Machine learning and its applications: A review', in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, Andhra Pradesh, India: IEEE, Mar. 2017, pp. 57–60. doi: 10.1109/ICBDACI.2017.8070809.
- [4] B. Yadranjiaghdam, N. Pool, and N. Tabrizi, 'A Survey on Real-Time Big Data Analytics: Applications and Tools', in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA: IEEE, Dec. 2016, pp. 404–409. doi: 10.1109/CSCI.2016.0083.
- [5] O. Obulesu, M. Mahendra, and M. ThrilokReddy, 'Machine Learning Techniques and Tools: A Survey', in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore: IEEE, Jul. 2018, pp. 605–611. doi: 10.1109/ICIRCA.2018.8597302.
- [6] M. Kaytan and I. B. Aydilek, 'A review on machine learning tools', in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya: IEEE, Sep. 2017, pp. 1–4. doi: 10.1109/IDAP.2017.8090257.
- [7] P. Saranya and P. Asha, 'Survey on Big Data Analytics in Health Care', in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Nov. 2019, pp. 46–51. doi: 10.1109/ICSSIT46314.2019.8987882.

## A Python Approach For Information Retrieval Using Vector Space Model (VSM)

**Narendra. M. Jathe,**

Assistant Professor

Smt. Narsamma, Arts Commerce & Science

College Amravati (India)

njathe@gmail.com

### **Abstract:**

The rapid growth of World Wide Web and the abundance of documents and different forms of information available on it, has recorded the need for good Information Retrieval technique. By and large, three classic framework models have been used in the process of retrieving information: Boolean, Vector Space and Probabilistic. Boolean model is a light weight model which matches the query with precise semantics. Because of its Boolean nature, results may be tides, missing partial matching, while on the contrary, vector space model, considering term-frequency, inverse document frequency measures, achieves utmost relevancy in retrieving documents in information retrieval.

Gerard Salton is often credited with developing the vector space model (VSM) for information retrieval (IR). The Vector Space Model is an algebraic model used for Information Retrieval. It represents natural language document in a formal manner by the use of vectors in a multi-dimensional space, and allows decisions to be made as to which documents are similar to each other and to the queries fired. vector space models which view documents and queries as vectors in a multidimensional vector space and use distance as a measure of similarity. The vector space model has been studied and used for some time by IR researchers and are quite well understood. This project implements and discusses the issues of information retrieval system with vector space model using PYTHON on College Website data of Amravati city domain.

**Keywords:** tf; idf; vector space model; cosine similarities; term-document; term-query matrices; dot products.

### **1. Introduction**

The amount of information from conventional databases and web sources has been growing exponentially in the past few years. Effective use of this huge amount of information has become a major challenge to the research communities working in this area. As far as the frequent changes in technologies and services are concerned, existing business concerns are tending towards making adaptive transformations so as to aggregate and integrate values from relevant information sources.

Information aggregation is a service that gathers relevant information from multiple sources to provide convenience and add value by analyzing the aggregated information for specific objectives using Internet technologies. The providers of this service are called "aggregators" in general. In a broader sense, information intermediaries such as newspapers, magazines, professional journals, and more recently, increasing number of web portals are information aggregators since they all collect information from multiple sources and disseminate it for convenient consumption [17].

Aggregation of information plays a vital role in the construction of knowledge based systems in various domains, ranging from healthcare, economics, applied science, artificial intelligence, and robotics to decision-making processes and machine learning.

Generally, user tends to search the information for a particular domain and as although the desired information is available on Web; it is in very scattered and varied structures across different geographic locations. Aggregating information in such a complex structure of Web is a challenging task. The combination of characteristics such as comparison, relation, intra-organization and inter-organization aggregation to maximize values in the aggregated information is really a challenging task. Although wrapping technology such as screen scraping, direct data feed with some encoding standard such as XML have been used by research groups in this field, there is a wide scope for research in information aggregation to facilitate varied aggregating services.

Colleges have made their public information available through their established web sites. As far as the contents of this particular domain is considered, most of the information is common in nature and visitors of this site are ranging from students, academicians, researchers to common people. But the information across these sites is very scattered and having wide scope to aggregate information in this domain and a very few research communities are working towards this domain for information aggregation. Hence the project proposed here is to build an efficient model of information aggregation from disparate source using web content mining with respect to Amravati city college websites.

Enormous amount of text material is increasing at exponential rate, especially with the increasing use and applications of Internet. Day by day it is becoming very difficult to retrieve the relevant information. Various approaches have been used by the researchers to get over the relevancy factor in information retrieval.

An information retrieval model is a quadruple consisting of document collection, set of queries, framework model and a ranking function associated with query-document. A framework model may be Boolean, vector space or probabilistic. Boolean model matches query with precise semantics in the document collection by Boolean operations with operators AND, OR, NOT. It predicts either relevancy or non-relevancy of each document, leading to the disadvantage of retrieving very few or very large documents. The Boolean model is the lightest model having inability of partial matching which leads to poor performance in retrieval of information. Vector space model is introduced by G. Salton in late 1960s in which partial matching is possible. Non-binary weights are used to weight the index terms in queries and in documents. These words are used for calculating degree of similarity between each document and the query. The ranked document set in the decreasing order of degree of similarity thus obtained is precise than the result of Boolean model. Index term weights can be calculated in many different ways.[18]

## **2. Related Work:**

Maron and Kuhns [4] in early 1960, described probabilistic indexing technique in a mechanized library system yielding probable relevance. Afterword in 1983, Salton and McGill wrote a book [1] which discusses thoroughly the three classic models in information retrieval namely, the boolean, the vector, and the probabilistic models. The book by van Rijsbergen [5] covers the discussion on three classic models and majority of the associated technology of retrieval system. Frakes and Baeza-Yates [6] edited the book on information retrieval which mainly deals with the data structures used in general information retrieval systems. Also, it includes the issue of relevance feedback as well as some query modification techniques [7] and Boolean operations and their implementations [8]. Verhoeff, Goffman, and Belzer [9] described the shortfall of boolean queries for information retrieval. The concept of using boolean formalism in other frameworks had been the great interest area of the researchers. Lee et al proposed a thesaurus-based boolean retrieval system for ranking [10]. Vector space model has been the most popular model in information retrieval among the research vicinity because of the research

outcome in indexing, term value specification in automatic indexing carried out by Salton and his associates [11, 12]. Most of this research deals with experiments in automatic document processing and different term weighting approaches for automatic retrieval [2, 13]. In 1972, Karen Sparck Jones introduced the concept of inverse document frequency, a measure of specificity [14, 15] and Salton and Yang uses it for automatic indexing to improve retrieval [12]. Raghavan and Wong [16] analyses vector space model critically with the conclusion that the vector space model is useful and which provides a formal framework for the information retrieval systems. Hongwei Zhu et al [17] in their paper presented dramatic growth in the amount of information on the Web. This trend will continue in the future as the last-mile bottleneck is being removed in developed countries and infrastructure is put in place in developing nations. Finding relevant information and extracting value from it is becoming more important for businesses and individuals. The emergence of information aggregation on the Internet provides an effective way of retrieving and managing relevant information that is dispersed all over the Web. The opportunities abound for businesses to provide value added services using aggregated information. A. B. Manwar et al [18] have develop Vector Space model for Information Aggregation, in which they mentioned Inter-document characterization and document frequency plays vital role in building ranks of the documents in vector space model. Xindong Wu et al [19] in their paper presented a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. Authors have analyzed the challenging issues in the data-driven model and also in the Big Data revolution.

### 3. Vector Space Model

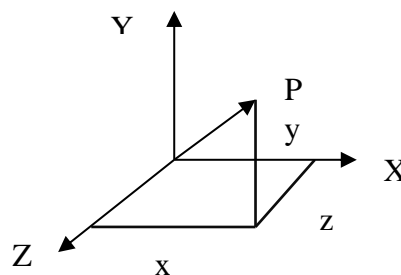
In the vector space model documents and queries are represented as vectors in a N-dimensional (multi-dimensional) hyperspace where each dimension corresponds to a possible document feature. Vector elements may be binary-valued, but they are generally taken to be weights that describe the degree to which the corresponding feature describes the document or query.

#### How it works

- ❖ Each document is broken down into a word frequency table
- ❖ The tables are called vectors and can be stored as arrays
- ❖ A vocabulary is built from all the words in all documents in the system
- ❖ Each document and user query is represented as a vector based against the vocabulary
- ❖ Calculating similarity measure
- ❖ Ranking the documents for relevance

#### 3.1 Vector

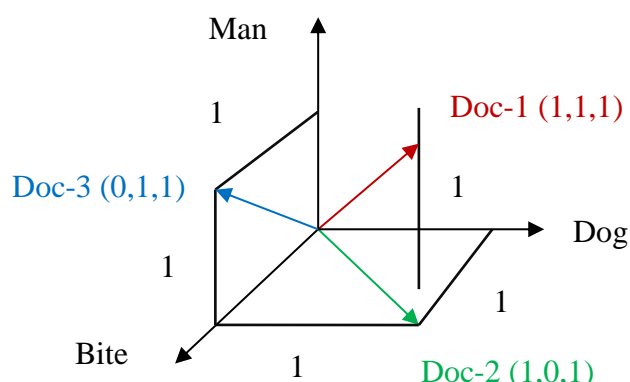
In mathematics, a vector is a point in a vector space and has length (from the origin to the point) and direction.



In the above figure  $X'$ ,  $Y'$ ,  $Z'$  are three dimensions creating three dimensional space, in which point P is a vector of elements  $x$ ,  $y$ ,  $z$ .

Similarly, each document and query represent as a vector in V-dimensional space. V denotes the size of the vocabulary.

	Dog	Man	Bite
Doc-1	1	1	1
Doc-2	1	0	1
Doc-3	0	1	1



### 3.2 Term Frequency

The term frequency  $tf(t,d)$  of terms in document is the number of times that terms 't' occurs in document 'd'.

### 3.3 Document Frequency (df):

Number of documents in which term 't' occurs.

### 3.4 Inverse-Document Frequency (idf):

- ❖ The inverse document frequency is a measure of how much information the word provides.
- ❖ A calculation designed to make rare words more important than common words.
- ❖ idf provides high values for rare words and low values for common words.
- ❖ idf calculates how much that term is common or rare across all documents. The idf, inverse document frequency for term 't', is given by

$$idf = \log((N+1)/df)$$

### 3.5 tf-idf:

The term which does not occur in multiple documents are rare terms (idf). Such rare term occurs frequently in documents (tf) have high importance (weight) in a document. The best known term-weighting schemes use weights which are given by

$$W_{i,j} = tf * idf$$

Such term-weighting strategies are called tf-idf schemes.

tf-idf increases with the number of occurrences within a doc. tf-idf increases with the rarity of the term across the whole corpus.

For the vector model, the weight  $W_{i,j}$  associated with a pair  $(K_i, d_j)$  is positive and non binary. Further, the index term in the query are also weighted. Let,  $W_{i,q}$  be the weight associated with the pair  $(K_i, q)$  where  $W_{i,q} \geq 0$ . Then, the query vector  $q$  is defined as  $q = (W_{1,q}, W_{2,q}, \dots, W_{t,q})$  where  $t$  is the total number of index terms in the system. The vector for a document  $d_j$  is represented by  $d = (W_{1,j}, W_{2,j}, \dots, W_{t,j})$ .

### 3.6 Length Normalization:

Automatic information retrieval, system have to deal with documents of varying lengths in a text collection. The document length normalization is used to fairly retrieve documents of all lengths.

Reasons for adopting a length normalization in VSM:

- ❖ Long documents have higher term frequency i.e. the same term appears more often.
- ❖ Long documents have more terms which increases the number of matches between the document and query.

Vectors can be normalized by simply converting each vector into a unit vector.

$$\text{Unit vector} = \frac{\vec{a}}{|\vec{a}|}$$

$$= \frac{(W_{1,j}, W_{2,j}, \dots, W_{t,j})}{\sqrt{\sum_i^t W_{i,j}^2}}$$

### 3.7 Cosine Similarity:

Cosine similarity measures the similarity between document vectors and query vector i.e. it finds relevant documents to a query. The vector model proposes to evaluate the degree of similarity of the document  $d_j$ , with regard to the query  $q$  as the correlation between the vectors  $d_j$  and  $q$ . the correlation can be measured by the cosine of angle between two vectors as,

$$\text{Similarity}(d, q) = \frac{\vec{d}}{|\vec{d}|} \cdot \frac{\vec{q}}{|\vec{q}|}$$

Where,  $|\vec{d}|$  and  $|\vec{q}|$  are the norms of the document and query vectors. The factor  $|\vec{q}|$  does not affect the ranking (i.e. the ordering of the documents) because it is the same for all document. The factor  $|\vec{d}|$  provides normalization in the space of the documents.

Since  $W_{i,j} \geq 0$  and  $W_{i,q} \geq 0$  varies from 0 to +1, the vector model ranks the documents according to their degree of similarity to the query. Documents might be retrieve even if it matches the query only partially.

## 4. Experimental Evaluation

### 4.1 Dataset for information retrieval system

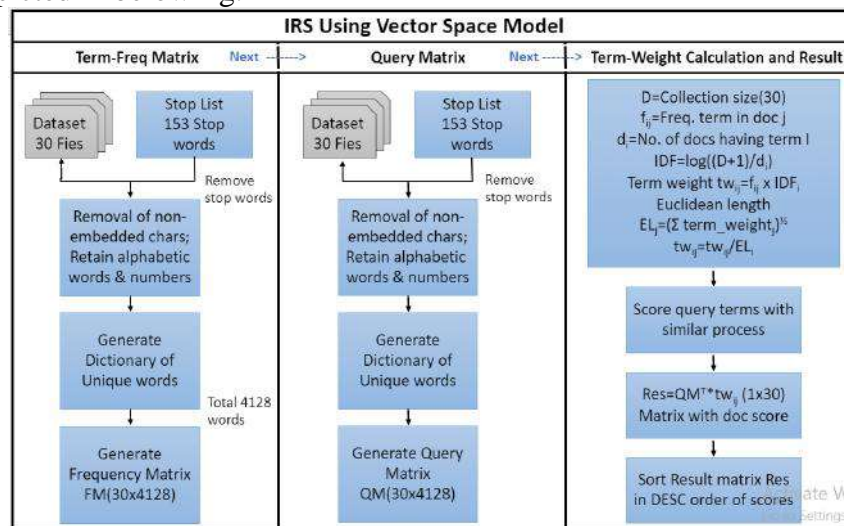
We use a website data of various streams of colleges of Amravati city domain. The collection contains total 30 different documents of different colleges. The details of different colleges from Amravati city with their respective streams like Arts, Commerce, Science, Animation, Engineering, Pharmacy, Medical, Architecture, Law, Agriculture etc.

### 4.2 Preprocessing

The compressed version of document text has been pre-processed to obtain a set of 30 individual abstract files. A special code has been written for this purpose in python.

### 4.3 Implementation

A Python is used to implement a vector space model for information retrieval; the complete process is depicted in below fig.



As shown in block diagram it consists of three stages:

#### 1. Generation of Term-Frequency matrix

It is term-frequency matrix of all unique terms in document  $d_j$  with  $1, 2, \dots, N$ . The term document matrix (FM) is  $M \times N$  matrix with  $t_i$  unique terms in dictionary ( $i=1, 2, \dots, M$ ) and  $N$  documents.

The elements of FM are represented as  $A_{i,j}$  which each element indicates the frequency of  $i$ th term in  $j$ th document.

All 30 documents is pre-processed to convert into individual 30text files. A SMART stop word list is used for the removal of stop words from the data collection of 30 files. Also, non-embedding special characters have been removed from these files. 22,502 words have been collected which are then processed to find the frequency of unique words in each documents. The dictionary of unique words is of 4128 words. Thus, the term frequency matrix is of size  $30 \times 4128$ .

## 2. Term-weight calculations and result

The result of experimentation is tabulated in below Table:

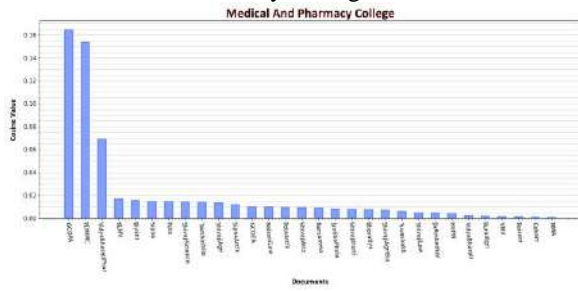
Rank	Q1		Q2		Q3		Q4		Q5	
	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.
1	5	0.1643	21	0.3963	20	0.3567	12	0.1731	24	0.2883
2	11	0.1536	15	0.3432	17	0.2684	6	0.1523	13	0.1368
3	27	0.0690	19	0.0771	18	0.1344	14	0.1035	23	0.0230
4	8	0.0170	29	0.0586	16	0.0756	23	0.0859	12	0.0134
5	1	0.0156	0	0.0549	22	0.0389	4	0.0565	0	0.0
6	23	0.0150	7	0.0547	12	0.0142	24	0.0440	1	0.0
7	12	0.0149	10	0.0470	1	0.0095	1	0.0316	2	0.0
8	22	0.0147	3	0.0178	19	0.0092	2	0.0277	3	0.0
9	25	0.0139	22	0.0173	13	0.0091	26	0.0260	4	0.0
10	17	0.0133	17	0.0140	10	0.0090	22	0.0173	5	0.0
11	24	0.0120	8	0.0135	7	0.0082	17	0.0140	6	0.0
12	4	0.0101	23	0.0112	0	0.0076	15	0.0125	7	0.0
13	15	0.0099	26	0.0105	29	0.0060	19	0.0119	8	0.0
14	13	0.0098	1	0.0096	23	0.0054	0	0.0107	9	0.0
15	19	0.0092	12	0.0084	21	0.0052	10	0.0097	10	0.0
16	10	0.0090	20	0.0082	3	0.0049	7	0.0094	11	0.0
17	7	0.0082	18	0.0081	11	0.0042	29	0.0087	14	0.0
18	20	0.0079	11	0.0060	6	0.0040	20	0.0082	15	0.0
19	0	0.0076	6	0.0050	8	0.0025	18	0.0081	16	0.0
20	18	0.0069	25	0.0037	5	0.0025	13	0.0070	17	0.0
21	29	0.0060	28	0.0035	26	0.0025	21	0.0062	18	0.0
22	21	0.0052	16	0.0035	27	0.0021	11	0.0060	19	0.0
23	3	0.0049	9	0.0034	28	0.0019	3	0.0057	20	0.0
24	6	0.0040	5	0.0027	14	0.0017	8	0.0055	21	0.0
25	26	0.0025	13	0.0027	25	0.0011	28	0.0035	22	0.0
26	16	0.0022	27	0.0023	2	0.0010	16	0.0035	25	0.0
27	28	0.0019	14	0.0022	9	0.0008	5	0.0027	26	0.0
28	14	0.0017	2	0.0018	4	0.0004	27	0.0023	27	0.0
29	2	0.0010	4	0.0006	15	0.0001	25	0.0014	28	0.0
30	9	0.0008	24	0.0004	24	0.0	9	0.0013	29	0.0

Rank	Q6		Q7		Q8		Q9		Q10	
	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.	Doc Id	Wt.
1	2	0.2725	0	0.1662	22	0.2442	7	0.3471	23	0.3472
2	22	0.0180	19	0.1658	20	0.2246	0	0.0	12	0.2497
3	17	0.0163	7	0.1518	19	0.1972	1	0.0	24	0.1782
4	1	0.0117	10	0.0946	18	0.1933	2	0.0	13	0.1614
5	19	0.0113	29	0.0905	17	0.1840	3	0.0	14	0.0706
6	10	0.0111	8	0.0897	28	0.0850	4	0.0	15	0.0515
7	7	0.0101	26	0.0839	21	0.0427	5	0.0	0	0.0
8	12	0.0097	3	0.0811	16	0.0370	6	0.0	1	0.0
9	20	0.0096	1	0.0696	11	0.0369	8	0.0	2	0.0
10	0	0.0093	22	0.0349	7	0.0349	9	0.0	3	0.0
11	18	0.0084	28	0.0227	0	0.0255	10	0.0	4	0.0
12	29	0.0074	9	0.0195	8	0.0123	11	0.0	5	0.0
13	23	0.0066	23	0.0169	25	0.0109	12	0.0	6	0.0
14	21	0.0063	25	0.0168	3	0.0093	13	0.0	7	0.0
15	3	0.0060	15	0.0060	27	0.0058	14	0.0	8	0.0
16	11	0.0054	12	0.0050	5	0.0048	15	0.0	9	0.0
17	6	0.0049	20	0.0050	12	0.0048	16	0.0	10	0.0
18	8	0.0031	16	0.0033	23	0.0041	17	0.0	11	0.0
19	5	0.0031	11	0.0033	1	0.0	18	0.0	16	0.0
20	26	0.0030	14	0.0025	2	0.0	19	0.0	17	0.0
21	13	0.0029	4	0.0019	4	0.0	20	0.0	18	0.0
22	16	0.0027	6	0.0016	6	0.0	21	0.0	19	0.0
23	27	0.0026	2	0.0	9	0.0	22	0.0	20	0.0
24	28	0.0023	5	0.0	10	0.0	23	0.0	21	0.0
25	14	0.0021	13	0.0	13	0.0	24	0.0	22	0.0
26	25	0.0014	17	0.0	14	0.0	25	0.0	25	0.0
27	9	0.0010	18	0.0	15	0.0	26	0.0	26	0.0
28	4	0.0005	21	0.0	24	0.0	27	0.0	27	0.0

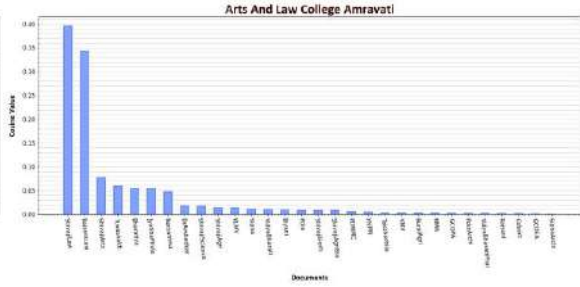
29	15	0.0002	24	0.0	26	0.0	28	0.0	28	0.0
30	24	0.0	27	0.0	29	0.0	29	0.0	29	0.0

**Query:**

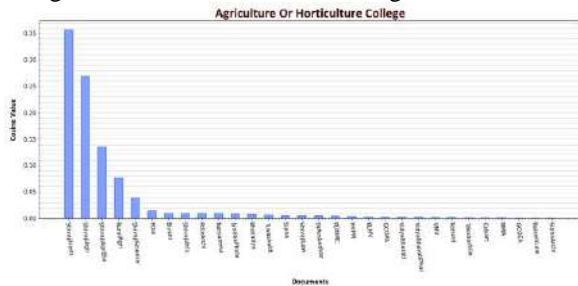
**1. Medical And Pharmacy College**



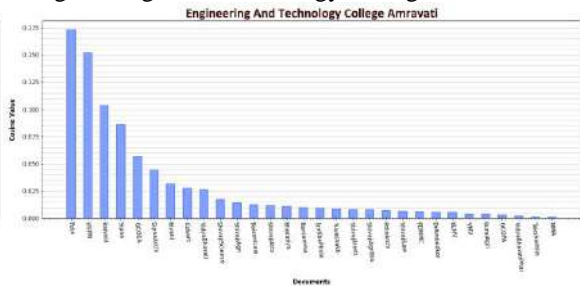
**2. Arts And Law College Amravati**



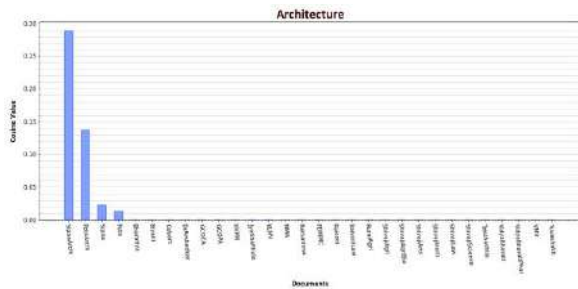
**3. Agriculture Or Horticulture College**



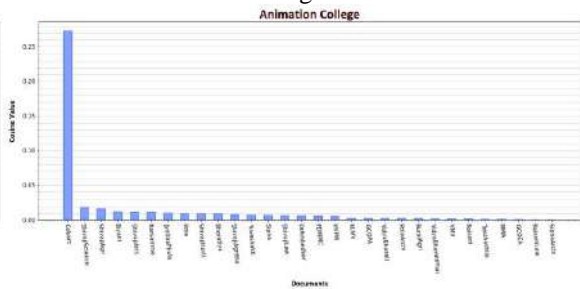
**4. Engineering And Technology College Amravati**



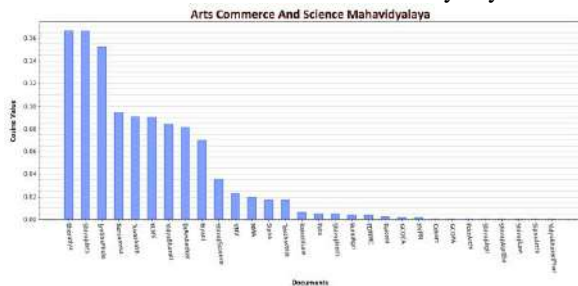
**5. Architecture**



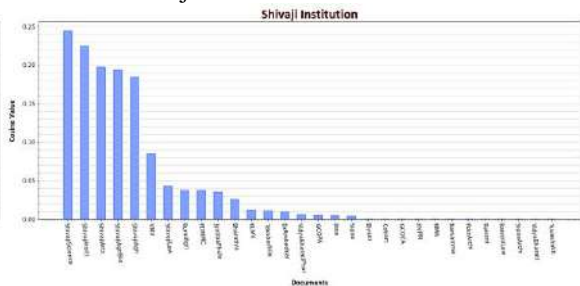
**6. Animation College**



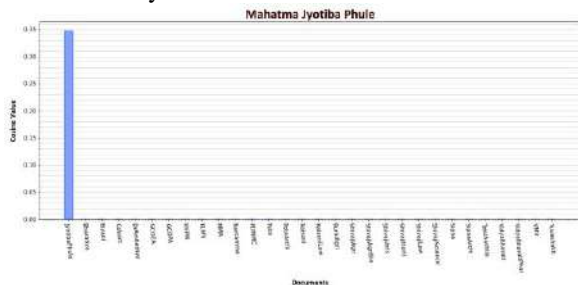
**7. Arts Commerce And Science Mahavidyalaya**



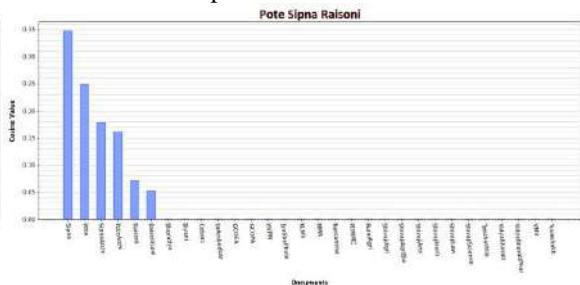
**8. Shivaji Institution**



**9. Mahatma Jyotiba Phule**



**10. Pote Sipna Raisoni**





## 5. Conclusion

The proposed framework will work to fetch information from different web sources to form a database. After collecting all the information, it will be processed by different web mining and data mining techniques and stored in the secondary database. The architecture design will be in terms of faster accessibility of information to the user.

The proposed framework will be capable of comparing the text-based information and help to represent more tabulated and systematic format. As the system is implemented with multiple data mining techniques, it will produce much better and accurate result.

The research work directs to develop web-based framework for knowledge representation using web mining and information retrieval techniques. This framework will be used by students, researchers, academicians and common people to present information using flexible user interface.

## References

- [1] G. Salton and M. J. McGill. Introduction to Modern Information Retrieval. McGraw-Hill Book Co., New York, 1983.
- [2] G. Salton and C. Buckley. Term-weighting approaches in automatic retrieval. *Information Processing and Management*, 24(5):513-523, 1988.
- [3] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, Introduction to Information Retrieval, Cambridge University Press, New York, USA, 2008.
- [4] M. E. Maron and J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Association for Computing Machinery*, 7(3):216-244, 1960.
- [5] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979.
- [6] W. B. Frakes and R. Baeza-Yates. *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.
- [7] D. Harman. Relevance feedback and other query modification techniques. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures and Algorithms*, pages 241-263. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.
- [8] S. Wartick. Boolean operations. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures and Algorithms*, pages 264-292. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.
- [9] J. Verhoeff, W. Goffmann, and Jack Belzer. Inefficiency of the use of Boolean functions for information retrieval systems. *Communications of the ACM*, 4(12):557-558, 594, December 1961.
- [10] J. H. Lee, W. Y. Kim, and Y. H. Lee. Ranking documents in thesaurus-based Boolean retrieval systems. *Information Processing and Management*, 30(1):79-91, 1993.
- [11] G. Salton and M. E. Lesk. Computer evaluation of indexing and text processing. *Journal of the ACM*, 15(1):8-36, January 1968.
- [12] Gerard Salton and C. S. Yang. On the specification of term values in automatic indexing. *Journal of Documentation*, 29:351-372, 1973.
- [13] G. Salton. *The SMART Retrieval System – Experiments in Automatic Document Processing*. Prentice Hall Inc., Englewood Cliffs, NJ, 1971.
- [14] K. Sparck Jones. A statistical interpretation of term specificity and its application to retrieval. *Journal of Documentation*, 28(1):11-20, 1972.
- [15] K. Sparck Jones. A statistical interpretation of term specificity and its application to retrieval. *Information Storage and Retrieval*, 9(11):619-633, 1973.
- [16] V. V. Raghavan and S. K. M. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Sciences*, 37(5):279-287, 1986.
- [17] Hongwei Zhu, Michael D. Siegel, Stuart E. Madnick "Information Aggregation - A Value-added E-Service". In the Proceedings of the International Conference on Technology, Policy, and Innovation, The Netherlands, 2001.
- [18] A. B. Manwar, Hemant S. Mahalle, K. D. Chinchkhede, Dr. Vinay Chavan " Vector Space Model For Information Retrieval: A Matlab Approach" *IJCSE*, ISSN : 0976-5166, Vol. 3 No. 2 Apr-May 2012, Pages 222-229
- [19] Xindong Wu ; Xingquan Zhu ; Gong-Qing Wu ; Wei Ding. "Data Mining with Big Data". *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, Issue 1, pages 97–107, 2014.

## 14

**Model-Driven Design of Graph Databases****Bhushan Jalamkar<sup>1</sup>, Dr.S.R.Thakare<sup>2</sup>**<sup>1</sup> Research Scholar, Department of Computer Science ,<sup>2</sup> Research Co-Guide, Associate Professor & Head Department of Computer Science & Application, Vidyabharti Mahavidyalaya, Amravati, Maharashtra, India**Abstract:-**

Database is not static but rapidly grows in size. These issues include how to allocate data, communication of the system, The coordination among the individual system, distributed transition control and query processing, concurrency control over distributed relation, design of global user interface, design of component system in different physical location, integration of existing database system. Design of database with Strategies using Various Methods. These Methods are concerned with solving the critical problems using Graph for improve the performance of databases.

**1. Introduction:-**

Design of centralized and distributed database is an alternative process that involves developing and refining a database structure based on the information and processing requirement of organization. Initially, a design strategy is significant for designing database. Basically, design strategies have two approaches, top- down and bottom-up.

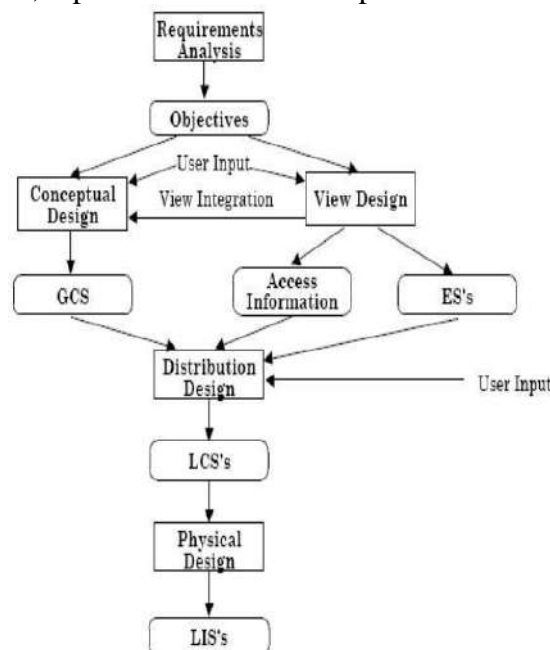


Figure 1: Top-down Design Process

**Top-down approach:-**Top-down design process is mostly used in designing system. The process starts from analysis of requirement including study of organisation, defining problems, defining objectives, conditions, scope and limitation. The next steps are conceptual design and view design as per requirement. Entity relationship model is used for conceptual design. The conceptual schema is designed through normalization. It creates abstract data structure to represent real facts. Defining user interface is on the basis of view design. All databases taken together in a distributed database are a virtual view on the basis of conceptual schema. The

entity and relationship requirement for all user views should be covered in conceptual schema. The conceptual schema uses defining global conceptual schema. View design activities are input of global conceptual schema and access information collection. Distribution objects need fragmentation and allocation over the system. These aspects use design local conceptual schema. The local conceptual schema maps the physical storage devices. The physical storage device carries out through physical design activity.

### **1.1 Bottom-up approach**

The bottom-up approach is convenient when the objective of the design is to integrate existing database system. The bottom-up design starts from local conceptual schema. The next step is gaining local schema into global conceptual schema.

The important aspect of design strategy is integrated multiple database system which is classified according to the autonomy, distribution and heterogeneity of the local systems. The logical data structure describes global conceptual schema (GCS). The global conceptual schema defines external schemas (ES) in distributed environment. The local internal schemas (LIS) represent physical data organisation on each site. These all are entities as shown in figure 1. is presented by local internal schema

## **1 Design Methods:-**

### **These Methods are as follows**

#### **2.1 Design Goal-oriented Schema:-**

Goal-oriented requirements analysis starts high-level goals, which are refined and interrelated to produce a goal model. The goal model captures not a single, but several alternative sets of data requirements, from which a particular one is chosen to generate the conceptual schema for the Biological database-to-be.

#### **Design OMT (Object Modeling Technique):-**

The OMT (Object Modeling Technique) method because it integrates a minimal set of concepts shared by several methods. These concepts are sufficient for our project. Moreover, OMT is based on various standards (Entity/ Association, Data flow diagram, State transition diagram) extended to the object paradigm. At last, the OMT graphical representation is expressive enough. A global conceptual schema of a DB managing a research centre documentation.

#### **2.2 Database Designer's Workbench:-**

The Database Designer's Workbench (or "Workbench") is a graphics oriented decision-support system to assist with the design of all aspects of a computerized database, from the initial specification of the system's requirements through its final physical structure. It provides a wide variety of design aids, or "tools," for designers to explore many design alternatives and to evaluate them precisely. These tools are presented in a homogeneous, graphically oriented environment that allows a designer to use familiar representations, store incomplete designs, progress smoothly from one design phase to the next, and iterate over previous design stages. The Workbench is therefore a real asset to the database design practitioner, seeking to improve productivity and the quality of design.

#### **2.3 Database Design Based on DFDC :-** Presents, The new method of

Database design, on the basis of Data Floating Direction

Chart (DFDC). The final design maps relation of database field. DFDC consists of tables and direction lines with formula mark, defined as 2-triples[80].  $DFDC = (T, L)$  where T is collection of tables, L is collection of lines. Table T defined as 2-triples,  $T = (N, I)$  where N is the name of tables and I is the set of the item. An item is denoted to be  $I = (IF, IN)$

where IF is a keyword flag for item, value for k represents keywords.

A line is denoted to be  $L = (ST.li, TT.Ij, F)$  Where  $ST.li$  is the source item meaning that the start of this line is the  $Li$  item in table  $ST.li$ .  $TT.Ij$  is the target item, that is, end of this line is the  $Ij$  item in the table  $TT$ , and  $F$  is a formula on the line.  $T$  is a table which contains set of items. It also uses functional dependency during design approach

**2.4 Database Design on based Star Model:-** Design Database, there are requirement centralized control of Database as well as Integration of system at that time specified the star model.

Star model of Database design is divided into three types: control data, integrated data, and security control.

□ **Control data:** by controlling the data, integrated system must provide to create data tables and to delete data tables. This model can not bring the impact of additions and deletions from Biological database.

□ **Integrated data:** this model must be fit to get together data from different government departments

□ **Secure access:** it is must be field-level security.

### 3. Proposed Methodology and its comparison:-

Its deals with the existing methodology and proposed methodology with respective biological database design, architecture supported to biological database, security model, transaction method and deadlock detection and recovery .

#### 3.1 Database Design Method:-

Design of centralized and distributed database is significant for further performance.

**3.2 Design Method with Work branch:-** Various approaches are used for design database. Work branch is one of the ways forthis. This way is graphically presented as well as through framework for a model. It is also called graphic-oriented decision, which includes five phases. This way is good for using large database, but it requires more time since complicity increases during designing.

#### 3.3 Design Method with goal oriented:-

This approach is goal-oriented. This concept is useful for analysis of the data as per requirement and to produce goal model. This goal model is a derived model of application domain. Lastly, conceptual scheme is created on basic domain. This is a way to produce absolute database but problem is created when further need arises to extend database. Hence it is not convenient for huge database.

#### 3.4 Design Method with object model technique:-

Another way is object model technique. This technique is used in conceptual design. It is based on various standards (entity / association, data flow diagram, state transition) or objects. The object model technique is perfect for distribution in a technical design. As such, the object linking is difficult to establish.

#### 3.5 Design Method with star model and data floating direction chart:-

One of the methods for design database is star model. This model provides security in terms of centralized control. This model is very useful in government sector and integrated system. This model is used only for secured database design. It is also critical for implementation. The new method is used for design database on the basis of data floating direction chart. This chart design methods, using table links properly. The limitation of this method is that it is only table-based and linking is critical except table-linking.

We have proposed design of database using graph. This method tries to solve problem related to other method. This method achieves scalability as well as dynamic change. It also achieves extend approach easily with accuracy. It requires less time for design. It also performs forward linking and backward linking correctly. Hence we have tried to maintain consistency with the help of this method.

#### 4.Design Issue for Proposed New Approach:-

##### 4.1 Requirement of New methodology

Many methods are used to design the centralized and distributed database to serve many aspects. The database designing is one of the methods in object modelling technique. This method shows various objects and it links with other data objects. Therefore, this method is used to design good database, but database should improve performance. Database designing uses objects molding technique with Graph. This method is called design of database using graph.

##### 4.2 Parameters for New Methodology:-

In the design of centralized and distributed database, we have considered aspects, viz: design methods, architecture support, security model, transaction, and deadlock detection and recovery, etc. We have proposed an data object related to databasedesign. This method is focused graphand data object are two parameters in this method. Graph represents flow of data towards proper direction. This Graph is based on object. First, we have to consider data objects and then design Graph. This is a way to use data objects and graph in this method.

##### 4.3 Proposed Methodology:-

###### Design of database using graphmethod:-

Database consists of data objects. Data objects are construction in the database on the basis of graph. Graph is made of nodes. Here Nodes consider as data objects

.Construction of algorithms for design of database on the base of graph. In this algorithms, give input as values of number of Data objects in form of Array. Two values are number of Data objects and Reference key . Number of Data objects and Reference key are represent N and R respectively. These numbers used within Array. Here put-up loop and perform task step by step,Call data object from the [N] sequencely.Create Reference Key of data object and stored in the [R].Selected data object and Adding into the Database. Repeat steps until last data object in array. Finally, get Graph as output.

##### 4.4Design of database using graph method:-

This Algorithm is basically focused on Data object,Reference key and graph. Graph is created by using algorithm as per requirement. First step is choosing the data objects. After this consider Reference keyof Data object which are classified in terms of table as per requirement. These Data object are connected to other data object using Reference key. Finally, Graph is created on basis of Data objects and Reference key .

The Algorithms as follows

```
// Graph G=(N,R)
// N is number of Data objects.
// R is number of Reference key.
//[N] is Array of data objects.
//[R] is Array of Reference key.
```

```
Input ( Number of data objects [N]) Output (Graph G )
Begin Loop
```

- 1) Call data object in [N].
- 2) Create Reference Key.
- 3) Include Reference Key into [R].
- 4) Select data object.
- 5) Add into database.

```
Repeat until last data object in Array.
Output(G); End;
```

**5. Conclusion:-**In the above proposed work, These are the various issues which are considered during the designing of the databases. These issues are concerned with solving the critical problems using Graph for improve the performance of databases. It is help to improve the performance as well as enhanced the quality of databases, and we can improve the clarity about the database.

**References:-**

- a. Richard E. COBB "The Databases Designers Workbench" Information System Research Group.(1984)
- b. Dixu "Distributed Database System Design" Minnesota State University Kankato.
- c. CLei Jiang , ThodorosTopaloglou, Alex Borgida, john Mylopoulos "Goal Oriented Conceptual Designing" (2007)
- d. Min CAO, Shu LUO, HuaikouMiao"Tree Based database desiging" School of Computer Engineering & Science 4<sup>th</sup> International Confernece on computer Science & Education.
- e. LubomyrSikora, uliyamiyushkovych "B+ tree in database Design for Decision making Information System" (2010).
- f. CHEN Bingchuan, LI Lei , Wang Heyong "New Methods of Database design Basd on DFDC" Second International Confernece on computer Science & Education (2010).
- g. .Wang heyong "Star like model database design for GEIS" Second International Confernece on E business & E government"

## Artificial Intelligence for Genetic Mutation Detection and its applications

**Bobade A D\*, Bhonde M M\*, Ingle S G\*\***

1. Department of Computer Science\*, 2. P. G. Department of Bioinformatics\*\*,  
Shri Shivaji Science, College, Amravati.

@corresponding author email: ashwinindbobade1@gmail.com\*, bioinfo.sham@gmail.com\*\*.

### Abstract:

Advancements in genomic sequencing technologies have provided unprecedented insights into the genetic basis of diseases, offering great potential for personalized medicine and early disease detection. However, the accurate identification of genetic mutations from vast and complex genomic datasets remains a challenging and labor-intensive task. This paper presents a comprehensive overview of Mutation Detection Using Artificial Intelligence (MDUAI), a burgeoning field at the intersection of genomics and artificial intelligence (AI). In this review, we delve into the fundamental principles of MDUAI and explore the diverse range of AI techniques employed for mutation detection. Significance of AI-driven mutation detection in various domains, including cancer research, rare disease diagnosis, and pharmacogenomics has been observed. Key machine learning algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and ensemble methods are highlighted for their efficacy in identifying mutations with high accuracy. Furthermore, the challenges and limitations associated with MDUAI, encompassing issues of data quality, interpretability, and ethical concerns. It insights into promising future directions, including the integration of multi-omics data, transfer learning, and the development of explainable AI models to enhance clinical adoption. By surveying the current landscape of MDUAI, this paper aims to facilitate a deeper understanding of the transformative potential of artificial intelligence in revolutionizing mutation detection. With AI-driven tools poised to play a pivotal role in advancing precision medicine, this review underscores the importance of continued research and collaboration between the genomics and AI communities to harness the full potential of MDUAI in improving healthcare outcomes.

**Keyword:** Artificial Intelligence, Pharmacogenomics, Mutations, Diagnosis, Research.

### 1. Introduction

Genetics and genomics have emerged as cornerstones of modern biology and medicine, unraveling the intricate genetic code that underpins the functioning of all living organisms, including humans. Understanding genetic mutations is pivotal in deciphering the genetic basis of diseases, predicting disease risk, developing targeted therapies, and advancing personalized medicine. Mutations, variations in the DNA sequence, can have profound effects on an individual's health, making their accurate detection and interpretation of paramount importance.

### Relevance of Artificial Intelligence (AI) and Machine Learning (ML) in Mutation Detection:

The sheer complexity and volume of genomic data generated by high-throughput sequencing technologies have presented a formidable challenge in the field of genetics and genomics. Traditional methods for mutation detection are often labor-intensive, time-consuming, and prone to human error. This is where AI and ML step in as transformative tools.

AI and ML algorithms possess the capability to efficiently analyze massive datasets, recognize subtle patterns, and make predictions with high accuracy[1]. When applied to mutation detection, they can expedite the identification of genetic variations, including single nucleotide polymorphisms (SNPs), insertions, deletions, and structural variations, even in the presence of noise and confounding factors. Techniques like deep learning, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and ensemble methods have demonstrated exceptional performance in mutation detection tasks[2]. The integration of AI and ML into genetics and genomics not only accelerates the research process but also enhances the precision and reliability of mutation identification. This synergy between AI/ML and genetics has the potential to revolutionize healthcare by enabling earlier disease diagnosis, facilitating drug discovery, and enabling personalized treatment strategies.

## 2. Objectives:

This paper aims to provide a comprehensive overview of the critical role of AI and ML in mutation detection within the context of genetics and genomics. To achieve this, the paper will address the following objectives:

1. To study the significance of mutation detection in genetics and genomics, emphasizing its implications for disease understanding, prevention, and treatment.
2. To study how AI and ML can be effectively applied to mutation detection.
3. To explore the diverse applications of AI-driven mutation detection, including cancer research, rare disease diagnosis, pharmacogenomics, and evolutionary biology, showcasing real-world examples of its impact.
4. To find limitations associated with AI/ML-based mutation detection.

Through this comprehensive exploration, the paper aims to shed light on the transformative potential of AI and ML in mutation detection, fostering a deeper understanding of their pivotal role in advancing genetics, genomics, and healthcare as a whole.

## 3. Background

### 3.1 Genetic Mutations:

Genetic mutations are alterations in the DNA sequence that can lead to changes in the encoded proteins or regulatory elements[3]. They are fundamental to the genetic diversity of all living organisms and play pivotal roles in various biological processes. Mutations can be categorized into several types:

- **Single Nucleotide Polymorphisms (SNPs):** These are the most common type of genetic variation, involving a change in a single nucleotide base pair. SNPs can influence susceptibility to diseases and response to medications.
- **Insertions and Deletions (Indels):** These mutations involve the insertion or deletion of one or more nucleotide bases in the DNA sequence. They can cause frameshifts and disrupt protein coding sequences.
- **Structural Variations:** These include larger-scale changes, such as duplications, inversions, translocations, and chromosomal rearrangements. Structural variations can have profound effects on gene regulation and function.

### 3.2 Significance in Various Fields:

- **Medicine:** Genetic mutations are central to understanding inherited diseases, predicting disease risk, and personalizing treatment plans. For example, mutations in the BRCA1 and BRCA2 genes are associated with increased risk of breast and ovarian cancers, informing decisions about screening and preventive measures.



- **Evolution:** Mutations are the driving force of evolution. They create genetic diversity within populations, allowing for adaptation to changing environments over generations.
- **Cancer Research:** Cancer often arises from the accumulation of mutations that disrupt normal cellular processes. Identifying these mutations is critical for developing targeted therapies and understanding the mechanisms of cancer progression.

### 3.3 Traditional Methods of Mutation Detection:

Traditionally, the detection of genetic mutations has relied on techniques such as:

- **Sanger Sequencing:** This method involves determining the order of nucleotide bases in a DNA strand. It is accurate but time-consuming and expensive, limiting its scalability[4].
- **Polymerase Chain Reaction (PCR):** PCR amplifies specific DNA regions, making them easier to analyse. However, it is limited to known mutations and may miss novel variants. Restriction Fragment Length Polymorphism (RFLP) Analysis: RFLP identifies mutations by cutting DNA at specific recognition sites. It is effective for some mutations but not all.
- **Next-Generation Sequencing (NGS):** It also known as high-throughput sequencing, has revolutionized the field of genomics and plays a pivotal role in mutation detection.

Traditional mutation detection methods suffer from several limitations:

1. **Labor-Intensive:** They are often labor-intensive, requiring manual sample preparation and analysis.
2. **Costly:** Many traditional methods are expensive, making large-scale screening and research challenging.
3. **Limited Scalability:** They may not be suitable for analyzing large genomic datasets or identifying rare mutations.

## 4. Introduction to AI and ML in Genomics

Artificial Intelligence (AI) and Machine Learning (ML) are transformative technologies that have found applications in genomics, addressing the limitations of traditional methods. In the context of genomics:

- **AI:** AI encompasses the broader field of computer science focused on creating intelligent agents capable of learning and making decisions. Machine Learning (ML) is a subset of AI that focuses on training algorithms to learn patterns from data.
- **Genomic Data:** Genomic data is vast and complex, making it an ideal candidate for AI/ML applications. These technologies can analyze large datasets, discover hidden patterns, and automate tasks like mutation detection.
- **Deep Learning:** Deep learning, a subset of ML, has been particularly effective in genomics. Deep neural networks, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), excel in tasks like variant calling and mutation classification.

AI and ML have the potential to revolutionize mutation detection by accelerating the process, improving accuracy, and enabling the analysis of massive genomic datasets, ultimately advancing our understanding of genetic mutations and their role in biology and medicine.

## 5. AI and ML Techniques for Mutation Detection

AI and ML techniques have been increasingly applied to mutation detection in genomics due to their ability to handle large and complex datasets. Some of the commonly employed techniques include:

1. **Convolutional Neural Networks (CNNs):** CNNs are well-suited for image and sequence data analysis. In mutation detection, they are often used to recognize patterns in DNA sequences, identifying regions that may contain mutations.
2. **Recurrent Neural Networks (RNNs):** RNNs are employed for sequence data analysis, including DNA and RNA sequences. They are useful for tasks like variant calling and predicting the impact of mutations on protein function.
3. **Random Forests and Decision Trees:** These ensemble methods are effective for classification tasks, such as distinguishing between benign and pathogenic mutations. They can handle a wide range of features derived from genomic data.
4. **Support Vector Machines (SVMs):** SVMs are used for binary classification problems in mutation detection, where the goal is to classify mutations as pathogenic or benign based on various features.

### 5.1 Data Sources and Preprocessing of genomic data:

Data sources for mutation detection encompass a variety of genomic and clinical data:

1. **Genomic Sequences:** Raw DNA or RNA sequences obtained from high-throughput sequencing technologies or already published NGS database repositories like NCBI SRA, European Nucleotide Archive (ENA) and UCSC are the primary source of data. This data may include whole-genome sequencing (WGS), whole-exome sequencing (WES), or targeted sequencing data[10-11].
2. **Variant Databases:** Publicly available variant databases like dbSNP, ClinVar, and COSMIC provide valuable reference data for comparing and annotating detected mutations[12].
3. **Clinical Records:** In clinical contexts, patient medical records, including family histories and clinical phenotypes, are crucial for interpreting the significance of mutations.

### 5.2 Preprocessing of the genomic data for AI and ML analysis includes.

- **Quality Control:** Removing low-quality reads and sequences with sequencing artifacts to ensure data reliability.
- **Alignment:** Mapping raw sequencing reads to a reference genome to identify genomic positions.
- **Variant Calling:** Identifying differences (variants) between the individual's sequence and the reference genome.
- **Feature Extraction:** Extracting relevant features from the data, such as variant type, allele frequency, and functional annotations.

### 5.3 Training and Validation Procedures:

Training and validation are critical to building accurate and reliable mutation detection models.

**Data Splitting:** The dataset is divided into training, validation, and test sets. The training set is used to train the model, the validation set is used for hyperparameter tuning and model selection, and the test set is used to assess the model's performance.

**Cross-Validation:** K-fold cross-validation may be employed to ensure robustness. The dataset is divided into K subsets, and the model is trained and validated K times, with each subset serving as a validation set once.

**Performance Metrics:** Metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are used to evaluate model performance.

**Overfitting Prevention:** Techniques like dropout and regularization are applied to prevent overfitting, ensuring that the model generalizes well to unseen data.

By employing these methods and techniques, AI and ML models can effectively detect genetic mutations, classify their significance, and predict their functional impact, contributing to advancements in genomics and personalized medicine.

## 6. Case Studies and Applications

### 6.1 Clinical Genomics:

**Case Study:** The Genomic Diagnosis in Neurodevelopmental Disorders (GeDiND) project applies AI to identify genetic mutations associated with neurodevelopmental disorders. By analyzing whole-exome sequencing data from affected individuals and their families, AI algorithms can pinpoint pathogenic mutations, aiding in the diagnosis and counseling of patients.

**Successes:** AI-driven mutation detection in clinical genomics has significantly improved the speed and accuracy of genetic diagnoses, leading to more precise treatments and better patient outcomes. It has enabled the discovery of novel disease-associated genes, expanding our understanding of rare genetic disorders.

**Challenges:** Challenges include the need for extensive data sharing and standardization, as well as the ethical considerations surrounding the use of genetic information in clinical practice.

### 6.2 Cancer Research:

**Case Study :** In cancer research, AI is employed to identify somatic mutations in tumor genomes. The Cancer Genome Atlas (TCGA) project[5] uses machine learning algorithms to analyze large-scale genomic and clinical data, identifying mutations associated with various cancer types[8]. For example, the identification of specific mutations in the EGFR gene has guided the development of targeted therapies for lung cancer.

**Successes:** AI has accelerated the identification of driver mutations in cancer, enabling the development of targeted therapies and immunotherapies. It has also facilitated the discovery of potential drug targets and biomarkers for early cancer detection.

**Challenges:** Challenges include data heterogeneity, data privacy concerns, and the need for interpretable AI models to guide clinical decision-making.

## 7. Conclusion

Mutation Detection Using Artificial Intelligence (MDUAI) represents a transformative approach in genomics, offering unprecedented opportunities for personalized medicine and disease understanding. The integration of AI and machine learning techniques has shown remarkable success in expediting mutation detection, enabling more accurate diagnoses and

advancing our knowledge of genetic contributions to various diseases. However, as the field continues to evolve, addressing challenges related to data quality, interpretability, and ethical considerations is paramount. Collaborative efforts between the genomics and AI communities, alongside ongoing research, will be crucial in harnessing the full potential of MDUAI. By embracing future directions such as integrating multi-omics data, implementing transfer learning, and developing explainable AI models, MDUAI can further contribute to revolutionizing healthcare outcomes. As AI-driven tools become integral to precision medicine, continuous exploration and refinement of MDUAI will play a pivotal role in shaping the future of genetics, genomics, and personalized healthcare.

## 8. References

1. Stewart, J., Sprivulis, P., & Dwivedi, G. (2018). Artificial intelligence and machine learning in emergency medicine. *Emergency Medicine Australasia*, 30(6), 870-874.
2. Shamsirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A. T., & Alinejad-Rokny, H. (2021). A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *Journal of Biomedical Informatics*, 113, 103627.
3. Sinclair, A. (2002). Genetics 101: detecting mutations in human genes. *Cmaj*, 167(3), 275-279.
4. Cotton, R. G. H. (1993). Current methods of mutation detection. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 285(1), 125-144.
5. Quazi, S. (2022). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology*, 39(8), 120.
6. Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia*, 2015(1), 68-77.
7. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25, 1315-1360.
8. Gupta, S., & Kumar, Y. (2022). Cancer prognosis using artificial intelligence-based techniques. *SN Computer Science*, 3, 1-8.
9. Ashrafuzzaman, M. (2021). Artificial intelligence, machine learning and deep learning in ion channel bioinformatics. *Membranes*, 11(9), 672.
10. Kodama, Y., Shumway, M., & Leinonen, R. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic acids research*, 40(D1), D54-D56.
11. Mackenzie, A., McNally, R., Mills, R., & Sharples, S. (2016). Post-archival genomics and the bulk logistics of DNA sequences. *BioSocieties*, 11, 82-105.
12. Senadheera, S. M., & Weerasinghe, A. R. (2020, November). Genomic Data Analyzing Workflow for Single Nucleotide Polymorphisms in Human Nervous System Cancers. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 107-112). IEEE.

## 16

**A Research on Cloud Computing****Miss Bhavana K. Bhagat**

Email: bhavanabhagat662@gmail.com

S. S. S. K. R. Innani Mahavidyalaya, Karanja Lad

**Miss. Pallavi A. Netankar**

Email: pallavinetankar@gmail.com

S. S. S. K. R. Innani Mahavidyalaya, Karanja Lad

**Miss. Komal S. Gupta**

Email: komalgupta2821998@gmail.com

S. S. S. K. R. Innani Mahavidyalaya, Karanja Lad

**Abstract:** Cloud computing is one of the best techniques for managing and allocating a lot of information and resources. Cloud computing is emerging rapidly and no doubt it is the next generation technology where humans will be using anywhere and anytime. In this internet world cloud computing is raising high by providing everything incense the required resources, application, software, hardware, computing power to computing infrastructure. business process to control collaboration. The IT team can adapt resources to changing and erratic requirements thanks to cloud computing. There is proof that cloud computing has a role in everyday life thanks to various applications in various contexts. This essay will cover every aspect of cloud computing, including its architecture, traits, types, service models, advantages, and challenges.

**Keywords:** Cloud computing, Architecture, Types, Benefits and Challenges

**I. Introduction:**

Cloud Computing is the delivery of computing services such as servers, storage, databases, networking, software, analytics, intelligence, and more, over the Cloud (Internet). Cloud Computing provides an alternative to the on-premises datacentre. With an on-premises datacentre, we have to manage everything, such as purchasing and installing hardware, virtualization, installing the operating system, and any other required applications, setting up the network, configuring the firewall, and setting up storage for data. After doing all the set-up, we become responsible for maintaining it through its entire lifecycle .Cloud computing is as a computing model in which massively scalable IT-enabled capabilities are offered as a service to numerous customers. It is the use of internet-based computer technology for a variety of services (as storage capacity, processing power, business applications, or components).. It is a set of network-enabled services that offer scalable, guaranteed, typically customized, relatively affordable services in an easy-to-use manner. Cloud computing is defined as a computing approach in which enormously scalable IT-related capabilities are delivered as a service through the internet to various.

**II. HISTORY, LITERATURE REVIEW AND METHODOLOGY ON CLOUD COMPUTING:****A. History of cloud computing**

Prof. John McCarthy (September 4, 1927 – October 24, 2011), an American computer scientist and professor at the Stanford University, popularized the concept of time sharing. He introduced the concept of time sharing at the MIT centennial. This idea of computer or Information utility

was very popular in the late 1960. IBM Digital Equipment Corporation and other mainframe providers work on this concept in the 1960-70.

## **B. Literature review on cloud computing**

As the years go on, the evolution of cloud computing started around 2006 when Amazon introduced "Amazon web services (AWS)" as an elastic cloud computing. Since the 2000 in the primary forms of cloud computing IaaS, PaaS and SaaS were formalized. Cloud computing was adapted by the SMB. Amazon started offering its Infrastructure of Application Service Provider. In 2008, Google introduced the "Beta" version of search engine. In 2012, Oracle rolled out Oracle cloud computing. This spacious memory allowed organizations to store, examine, and acquire valuable data on customers' information, interests, and reaction

## **C. Methodology**

According to [34], the research technique approaches and methods form the research methodology used in collecting information regarding different parts of a problem. The method adopted is descriptive method. Materials for the writing was sourced through Google scholar, Scopus. We came across different views about cloud computing by different authors. Different authors had their own definition of cloud computing. The key thing we discovered about cloud computing is that you pay as you use the service.

## **III. CLOUD COMPUTING ARCHITECTURE:**

A cloud customer is a collection of computer hardware and software that leverages cloud computing to offer applications designed specifically to deliver cloud services.

The following three service models are used to categorize cloud services.

### **A. Infrastructure-as-a-service (IaaS):**

In this model, You can either use servers or storage in the cloud. In this model, you do not have to purchase and maintain your own IT hardware. However, you need to install your application on your cloud based hardware resources. Contrary to that model, this tactic is different. Google, App Engine, Microsoft Azure, Java, and developer tools are a few examples of infrastructure as a service.

### **B. Platform as a service (PaaS):**

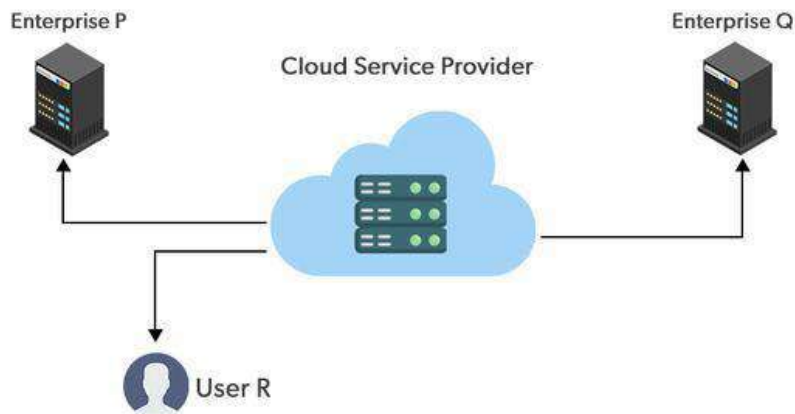
In this model you can use the cloud as a platform to develop and sell software application. As a result, PaaS offers development infrastructure such as configuration management, tools, programming environments, and other components in addition to a hosting environment. Microsoft Azure, Google App Engine, developer tools, and Java are a few examples of PaaS.

### **C. Software as a service (SaaS)**

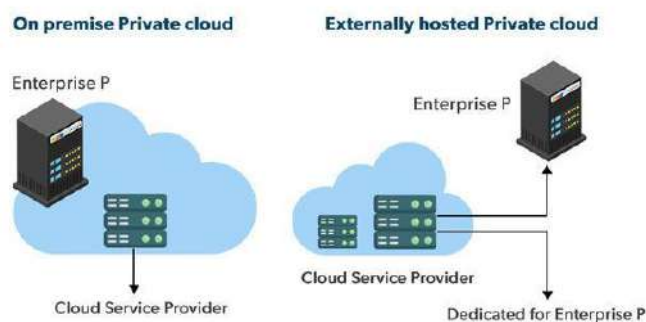
In this model you can use various software application such as CRM and ERP, and collaboration tools on the web. To achieve economies of scale and optimization in terms of speed, availability, disaster recovery, maintenance, and security, users of the applications of various cloud consumers are grouped on the SaaS cloud in a single logical environment

### III. TYPES OF CLOUD COMPUTING

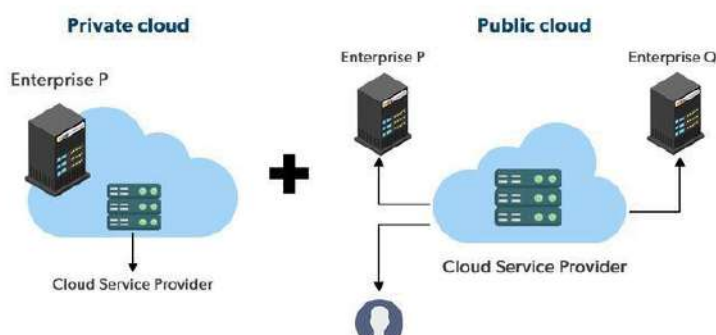
**Public Cloud:** The public cloud is a computing service supplied by the third party providers atop the public internet . These services are available for any user who wants to use them and they have to pay only for the services they consumed.



**Private Cloud:** The computing services provided over the internet or private network come under the private cloud and these services are offered only to the selected users in place of common people . A higher security and privacy is delegated by private clouds through the firewall and internal hosting .



**Hybrid Cloud:** Hybrid cloud is the combination of public cloud and private cloud. In the hybrid cloud, each cloud can be managed independently but data and applications can be shared among the clouds in the hybrid cloud .



#### IV. BENEFITS OF CLOUD COMPUTING

**Cost Saving:** In cloud computing users have to only pay for the services they consumed. Maintenance cost is low as user do not need to purchase the infrastructure

**Flexibility:** Cloud computing is scalable. The rapid scale up and down in the operations of your business may require quick adjustment of hardware and resources so in order to manage this variations cloud computing provide flexibility.

**Enhanced Security:** Cloud computing provide high security by using the data encryption, strong access controls, key management, and security intelligence.

#### CLOUD COMPUTING CHALLENGES

##### A. Security and privacy

These include the organizational and technical difficulties of maintaining a sufficient level of data privacy and security in cloud services.

##### B. Data Management

Due to the increased data-intensive applications that cloud computing enables at the largest scale, there is an increased need for efficient data management solutions. Data storage falls under this heading. Data segmentation, recovery, location, authenticity, anonymization, and backup are all parts of data.

##### C. Economic Challenges

The cost of the physical infrastructure and the administrative costs connected to it are essential in determining the viability of the business from an economic perspective.

#### CONCLUSION

In this research, we discussed the architecture, types, characteristics of cloud computing is key in information technology as it reduces cost for organizations and makes it easier to access files. It also helps to reduce data delay and redundancy. Any organization that wants to adopt cloud computing should consider the key challenges which is security and privacy.

#### REFERENCES

- [1] M. Bamiah, S. Brohi, S. Chuprat, and M. N. Brohi, "Cloud implementation security challenges," 2012
- [2] I. Baldini, P. Castro, K. Chang, P. Cheng, S. Fink, V. Ishakian, N. Mitchell, V. Muthusamy, R. Rabbah, A. Slominski, and P. Suter, "Serverless Computing: Current Trends and open problems," Research Advances in Cloud Computing, pp. 1–20, 2017.
- [3] "Cloud Computing: Concept, Technology and Architecture" Thomos Erl et.al., Pearson, 2013.
- [4] "Virtualization Essentials", Matthew Portnoy, sybx 2012.
- [5] VeritisAdmin, "Cloud computing trends, Challenges & Benefits," Go to Veritis Group Inc., 12-Oct-2022. [Online]. Available: <https://www.veritis.com/blog/cloud-computingtrends-challenges-and-benefits/>. [Accessed: 15Dec-2022].
- [6] N. Robinson, L. Valeri, J. Cave, T. Starkey, H. Graux, S. Creese and P. Hopkins, "The
- [7] CloudUnderstanding the Security, Privacy and Trust Challenges", RAND Corporation, 2011.
- [8] Dan Morill, "Cloud Computing in Education", <http://www.cloudave.com/14857/cloudcomputingin-education/2011>.
- [9] S. A. Sheik and A. P. Muniyandi, "Secure authentication schemes in cloud computing with glimpse of Artificial Neural Networks: A Review," Cyber Security and Applications, vol. 1, p. 100002, 2023.



## Impact of Hadoop Technology in Cloud Computing

**Prof. Sameena A.Kazi**

M.C. A, M.C.M, Dept of Management  
Vidya Bharati Mahavidyalaya, Amravati  
Email: [sakazi140@gmail.com](mailto:sakazi140@gmail.com)

### Abstract:

The concept of Cloud Computing has been distinguished as one of the major computing models in recent years. Cloud computing has become a great innovation that has important consequences not just for services on the internet but also for the entire Information technology (IT) market. It is an internet-based technology that enables small business and organizations to use highly sophisticated computer applications. Basically, it refers to the applications and services offered over the Internet. These services are offered from data centers all over the world, which collectively are referred to as the "cloud". Despite of its advantages, Cloud Computing has many drawbacks (low scalability, no support for stream data processing). Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs. Hadoop framework needs to be implemented in cloud computing to overcome its drawbacks. This paper highlights the role of Hadoop in Cloud Computing environment.

**Keywords** - Cloud Computing, Hadoop, HDFS, MapReduce

### 1. INTRODUCTION

In today's environment almost all the companies have migrated their applications and data to the cloud due to the popularity of the Internet along with a sharp increase in the network bandwidth. This has resulted in the generation of huge amounts of data on regular basis. So, in current circumstances, management of big distributed data like cloud is a big challenge. For processing such gigantic amount of data, the traditional methods of database management are not appropriate since these approaches fail to handle such size of data. Hence, in order to handle such huge volume of heterogeneous data, companies are now coming up with different alternatives. One of the widely accepted solutions is Hadoop, which has attracted great attention due to its fault tolerance, highly available, accessible, scalable, ease of programming, have huge flexible storage and low-cost features. Hadoop is an open-source software framework for storing and processing big data. It was created by Apache Software Foundation in 2006, based on a white paper written by Google in 2003 that described the Google File System (GFS) and the MapReduce programming model. The Hadoop framework allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. It is used by many organizations, including Yahoo, Facebook, and IBM, for a variety of purposes such as data warehousing, log processing, and research. Hadoop has been widely adopted in the industry and has become a key technology for big data processing.

## 2. HADOOP FRAMEWORK

Hadoop is an open-source framework that allows users to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Hadoop is an Apache open-source framework written in Java that allows the distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation, and storage.

Hadoop provides two things: Storage and Compute. In Hadoop, storage is provided by Hadoop Distributed File System (HDFS), and compute is provided by MapReduce. Hadoop is an open-source implementation of Google's distributed computing framework, which is a proprietary framework. It consists of two parts: HDFS, which is modelled after Google's GFS, and Hadoop MapReduce, which is modelled after Google's MapReduce.

MapReduce is a programming framework, which organizes multiple computers in a cluster in order to perform the calculations needed. It takes care of distributing the work between computers and of putting together the results of each computer's computation. Just as importantly, it takes care of hardware and network failures so that they do not affect the flow of computation. It is required that a problem has to be broken into separate pieces which can be processed in parallel by multiple machines.

HDFS gives the programmer unlimited storage; however, there are additional advantages of HDFS which are listed below:

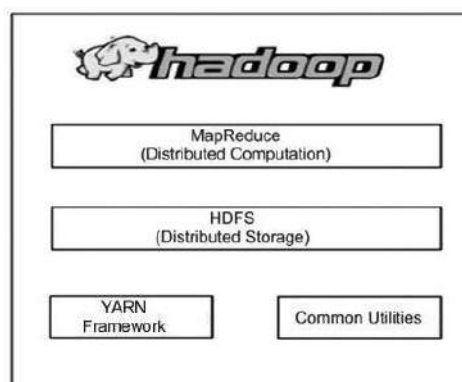
**i. Horizontal scalability:** Thousands of servers hold petabytes of data. When you need even more storage, you don't switch to more expensive solutions but add servers instead.

**ii. Commodity hardware:** HDFS is designed with relatively cheap commodity hardware in mind. HDFS is self-healing and replicating.

**iii. Fault tolerance:** Every member of the Hadoop knows how to deal with hardware failures. If there are 10 thousand servers, then one server will fail every day on average. HDFS foresees that by replicating the data, by default three times, on different Datanode servers. Thus, if one data node fails, the other two can be used to restore the third one in a different place.

## 3. HADOOP ARCHITECTURE

Hadoop architecture has two major layers: Processing/Computation layer (MapReduce), and Storage layer (Hadoop Distributed File System) as shown in Figure 1.



**FIGURE 1.** The Hadoop architecture

### **i. MapReduce**

MapReduce is a parallel programming model for writing distributed applications devised at Google for efficient processing of large amounts of data (multi-terabyte datasets), on large clusters (thousands of nodes) of commodity hardware in a reliable, and fault-tolerant manner. It is discussed in detail in previous sections. The MapReduce program runs on Hadoop, which is an Apache open-source framework.

### **ii. Hadoop Workflow**

It is quite expensive to build bigger servers with heavy configurations that can handle large-scale processing, but as an alternative, you can tie together many commodity computers with single CPUs, as a single functional distributed system and, practically, the clustered machines can read the dataset in parallel and provide a much higher throughput. Moreover, it is cheaper than one high-end server. So this is the first motivational factor behind using Hadoop that it runs across clustered and low-cost machines.

Hadoop runs code across a cluster of computers. This process includes the following core tasks that Hadoop performs:

- Data is initially divided into directories and files, files are further sub divided into uniform sized blocks.
- These files are then distributed across various cluster nodes for further processing.
- HDFS, being on top of the local file system, supervises the processing.
- Blocks are replicated for handling hardware failure.
- Checks that the code was executed successfully.
- Sending the sorted data to a certain computer.
- Writing the debugging logs for each job.

## **4. HADOOP DISTRIBUTED FILE SYSTEM (HDFS)**

The HDFS is based on a distributed file system that is designed to run on commodity hardware. It has many similarities with the existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in a redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available for parallel processing. HDFS can be deployed on low cost hardware and is highly fault tolerant. It provides high throughput access to application data and is suitable for applications having large datasets.

HDFS is designed for storing very large files with write once and read many times capability. This is the main difference between HDFS and a generic file system. A generic file system allows files to be modified. HDFS is not a good fit for low latency data access when there are lots of small files and for modification at arbitrary offset in the file. Files in HDFS are broken into block-sized chunks, the default size being 64 MB, which are stored as independent units. An HDFS cluster has two types of node operating in a master-worker pattern: a NameNode (master) and a number of DataNodes (workers). The Namenode manages the file system namespace. It maintains the file system tree and the metadata for all the files and directories in the tree. The Namenode also knows the Datanodes on which all the block for a given file are located. Datanodes store and retrieve blocks when they are told to and they report back to the Namenode periodically with lists of blocks that they are storing.

Apart from the above mentioned two core components, Hadoop framework also includes the following two modules:

**Hadoop Common:** These are Java libraries and utilities required by other Hadoop modules.

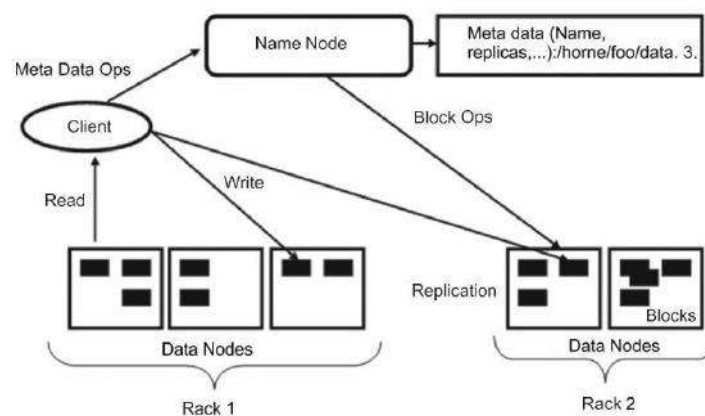
**Hadoop YARN:** This is a framework for job scheduling and cluster resource management.

#### 4.1 HDFS Features

- Best suited for the distributed storage and processing.
- Built-in servers of Namenode and Datanode help users to easily check the status of clusters.
- Hadoop provides a command interface to interact with HDFS.
- Streaming access to file system data.
- HDFS provides file permissions and authentication.

#### 4.2 HDFS Architecture

Figure 2. shows the architecture of a Hadoop File System.



**FIGURE 2.** HDFS architecture

HDFS follows the master-slave architecture and it has the following elements:

#### **Namenode:**

The Namenode is the commodity hardware that comprises of GNU/Linux operating system and the Namenode software. It is software that can be run on commodity hardware. The system having the Namenode acts as the master server and it does the following tasks:

- Manages the file system Namespace
- Regulates a client's access to files

- Executes file system operations such as renaming, closing and opening files, and directories

**Datanode:**

The Datanode is a commodity hardware having the GNU/Linux operating system and Datanode software. For every node (commodity hardware/system) in a cluster, there is a Datanode. These nodes manage the data storage of their systems.

- Read-write operations are performed on the file systems as per client requests.
- They also perform operations such as block creation, deletion, and replication according to the instructions of the Namenode.

**Block:**

Generally, the user data is stored in the files of HDFS. The file in a file system is divided into one or more segments and/or stored in individual data nodes. These file segments are called blocks. The minimum amount of data that HDFS can read or write is called a Block. A block default size is 64MB, but it can be increased according to need and make necessary changes in HDFS configuration.

## 5. CLOUD COMPUTING

Cloud computing (Mell et al. 2015) is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud is a network of servers that pools different resources. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. Cloud computing is a practical approach to experience direct cost benefits and it has the potential to transform a data center from a capital-intensive set up to a variable priced environment.

Cloud Providers offer services that can be grouped into categories depending upon either the type of service being provided or on the basis of location as shown in figure 3. According to Almorsy et al. (2016), the three basic service models are described as follows:

**1. Infrastructure-as-a-service (IaaS):** where cloud providers deliver computation resources, storage and network as an internet-based services. This service model is based on the virtualization technology. Amazon EC2 is the most familiar IaaS provider.

**2. Platform-as-a-service (PaaS):** where cloud providers deliver platforms, tools and other business services that enable customers to develop, deploy, and manage their own applications, without installing any of these platforms or support tools on their local machines. The PaaS model may be hosted on top of IaaS model or on top of the cloud infrastructures directly. Google Apps and Microsoft Windows Azure are the most known PaaS.

**3. Software-as-a-service (SaaS):** where cloud providers deliver applications hosted on the cloud infrastructure as internet-based service for end users, without requiring installing the applications on the customers' computers. This model may be hosted on top of PaaS, IaaS or directly hosted on cloud infrastructure. Salesforce, CRM is an example of the SaaS provider.

As per Zhang, (2010) enterprises can choose to deploy applications on Public, Private or Hybrid clouds. These are described as follows:

**1. Public clouds:** A cloud in which service providers offer their resources as services to the general public. Public clouds offer several key benefits to service providers, including no initial capital investment on infrastructure and shifting of risks to infrastructure providers.

**2. Private clouds:** Also known as internal cloud, and they are designed for exclusive use by a single organization. A private cloud may be built and managed by the organization or by external providers. A private cloud offers the highest degree of control over performance, reliability and security

**3. Hybrid clouds:** A hybrid cloud is a combination of public and private cloud models that tries to address the limitations of each approach. In a hybrid cloud, part of the service infrastructure runs in private clouds while the remaining part runs in public clouds. Hybrid clouds offer more flexibility than both public and private clouds.



**FIGURE 3.** Cloud Computing

## 6. CONCLUSION

This paper described a systematic flow of survey on Hadoop role in context of cloud computing. The key issues, including Hadoop architecture, HDFS, MapReduce and Cloud Computing environment is described. Hadoop is widely used for large scale data processing in cloud platforms. It is an open-source implementation of framework which hides the complexity of parallel execution across hundreds of servers in a cloud environment. It allows developers to process terabytes of data. After conducting a comprehensive review, this paper concludes that, on cloud platform, Apache Hadoop is a powerful open-source software platform. Hadoop's MapReduce and HDFS is used to deliver very high data availability and to analyze enormous amounts of information quickly. Hadoop has been proven to be a useful tool for distributing the processing over as many processors as possible. In Future, Hadoop will become the first choice for cloud computations.

**REFERENCES**

- [1] Alam, M., & Shakil, K. A. (2016). Big Data Analytics in Cloud environment using Hadoop. *arXiv preprint arXiv:1610.04572*.
- [2] Ikhlaq, S., & Keswani, B. (2016). Computation of Big Data in Hadoop and Cloud Environment. *IOSR Journal of Engineering*, 6(1), 31-39.
- [3] Patil, A. U., Patil, R. U., Pande, A. P., & Patil, B. S. Secured Hadoop as A Service Based on Infrastructure Cloud Computing Environment.
- [4] Ansari, S. M., Chepuri, S., & Wadhai, V. (2015). Efficient Map Reduce Model with Hadoop Framework for Data Processing.
- [5] Kaur, G., Kaur, M., 2015. Review Paper On Big Data Using Hadoop. *International Journal of Computer Engineering & Technology (IJCET)* 6, 65–71.
- [6] [www.greeksforgeeks.org/hadoop-on-introduction/](http://www.greeksforgeeks.org/hadoop-on-introduction/)
- [7] Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, 1(1), 7-18.

## Internet Of Things: Applications and Security Challenges

**Miss.Ankita Suresh Chambatkar** **Miss.Nikita Sakharam Dhande**

M.Sc 2yr  
Department of Computer Science  
Vidya Bharti Mahavidyalaya,  
Amravati

chambatkar2022ankita@gmail.com

M.Sc 2<sup>nd</sup> yr  
Department of Computer Science  
Vidya Bharti Mahavidyalaya,  
Amravati

nikitadhande1407@gmail.com

**Prof.S.M Mohod**

Department of Comp Science  
Vidyabharati Mahavidyalaya  
Amravati

shitalmmohod@gmail.com

**Abstract** - A detailed examination of several Internet of Things (IoT)-based applications is provided in this paper. It highlights that, as opposed to individuals, items are connected via the internet. The Internet of Things (IoT) is a massive network of interconnected objects and people that gather and exchange data about their usage and their surroundings. Applications for IOT can be found in industries like agriculture, healthcare, supply chain management, and defense. Finally, the article addresses IOT-related challenges. Despite its benefits, IOT has some serious problems, such as security and privacy. It also offers an analysis of the Internet of Things and its uses in various scientific and technological domains. A literature study is included with the introduction of the Internet of Things. The paper also covers the IoT's components and architecture, in addition to its various uses.

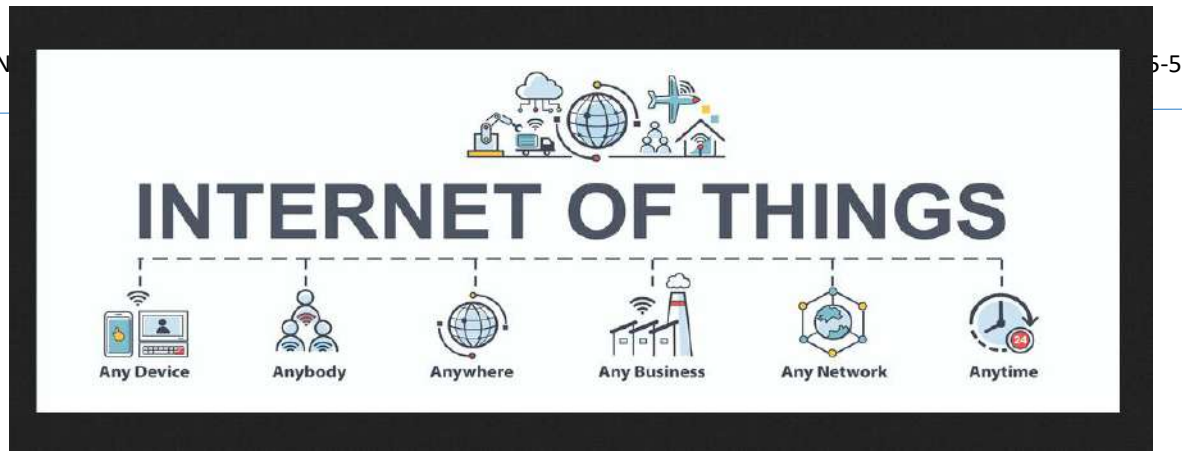
**Keywords** – Internet of Things (IoT), Information about Iot, Characteristics of IoT, Application of Iot, Key features of IoT.

### INTRODUCTION

The Internet of Things (IoT) has become a more viable platform for implementing this kind of innovative technologies. While cloud computing is not a novel concept in this industry, it has been utilized to symbolize the world of omnipresent computing. The seventh ITU Internet Report was first published in 1997 under the heading "Challenges to the Network." It was originally used in the RFID magazine in 1999 by Kevin Ashton. The term was modified to "Internet of things" in 2005. According to Kevin, the Internet of Things would allow networked devices to share information about real-world items via the internet. Most of the IoT architectures that have been suggested recently are utilized to broadcast information via social networks using web semantics.

**INFORMATION-** IOT, or the Internet of Things the internet of things, or IoT, is a network of connected computers, digital and mechanical devices, items, animals, and people that can transfer data over a network without requiring human-to-human or human-to-computer interaction. All of these devices are assigned unique identifiers (UIDs). Any natural or artificial object that can be given an IP address and be able to transfer data over a network, such as a person with an implanted heart monitor, a farm animal with a biochip transponder, an automobile with sensors to warn the driver when tire pressure is low, or any other combination of these, can be considered a thing in the internet of things[1][2].





## ARCHITECTURE OF IOT

The best architecture design serves as the cornerstone for creating a privileged Internet of Things system. It addressed numerous challenges related to scalability, routing, networking, and other aspects of the IoT environment. Generally speaking, the three primary dimensions of the IoT architecture are: All items connected to the Internet of Things (IoT) can be classified into three categories: (i) information items, which include sensing, identifying, and controlling items; (ii) independent networks, which have multiple features like self-configuration, self-protection, self-adaptation, and self-optimization; and (iii) intelligent applications, which have intelligent behavior over the Internet in general. These applications can be classified based on their intelligent behavior, which can include data processing, intelligent control, and intelligent exchange methods through network items. All IoT-related applications can be categorized based on these dimensions. Connectivity between the physical, digital, and social realms is the aim of the IoT's future architecture. [3] These units of IoT are designed to mimic human neural networks and offer solutions for particular applications. Space-time consistency, cyber, physical, and social co-existence, connectivity and interaction, and multi-identity status are the primary features of the U2IoT concept.

## CHARACTERISTICS OF INTERNET OF THINGS (IOT)

Some most popular characteristics of Internet of things are:

- Intelligence
- Connectivity
- Dynamic Nature
- Enormous scale
- Security

### Intelligence –

The IoT is intelligent because it combines hardware, software, and algorithms. In the context of the Internet of Things, ambient intelligence improves its skills to help objects respond intelligently to a given circumstance and assist them in completing certain tasks. Notwithstanding the widespread use of smart technologies, intelligence in the Internet of Things is limited to the means of communication between devices; graphical user interfaces and conventional input techniques are the ways in which users and devices interact.

### Connectivity-

The Internet of Things is made possible by connectivity, which connects commonplace objects. These things must be connected because even basic interactions between objects in an Internet of Things network can lead to collective intelligence. It makes the items compatible with and accessible over networks. The networking of smart objects and apps can open up new markets for the Internet of Things with this connectivity. In the Internet of Things, connectivity is more than just installing a Wi-Fi module and calling it a day. Network compatibility and accessibility are made possible by connectivity. Getting on a network is known as accessibility, whereas

sharing the capacity to create and consume data is known as compatibility. If this sounds familiar, it is because Metcalfe's Law applies to the Internet of Things.

#### **Dynamic Nature -**

The main function of the Internet of Things is data collection from its surroundings, which is made possible by the dynamic changes that occur in and around the devices. These devices' states vary on a dynamic basis; for instance, whether they are sleeping, waking up, connected, or not, and in what context—temperature, location, and speed, for example. The quantity of devices also varies dynamically with a person, place, and time, in addition to the device's status. The context of devices, which includes their location and speed, as well as their states—such as sleeping and waking up, connected or not—all fluctuate dynamically.

#### **Enormous scale -**

There will be a lot more gadgets than the ones currently connected to the Internet that need to be managed and communicate with one another. It becomes more important to handle the data produced by these devices and understand it for use in applications. In its estimation study, Gartner (2015) validates the massive scope of the Internet of Things, estimating that 5.5 million new items will be connected daily and that there will be 6.4 billion connected items in use globally in 2016, an increase of 30% from 2015. According to the research, there will be 20.8 billion linked devices by 2020. There will be at least an array of devices that require management and communication with one another.

#### **Security -**

Security risks are inherent to Internet of Things devices. It would be a mistake to ignore the security risks connected with the Internet of Things, as we enjoy increased productivity, new experiences, and other advantages from it. The IoT has a great deal of transparency and privacy concerns. Establishing a security paradigm is necessary in order to safeguard the networks, endpoints, and data that are moved between them [4].

### **APPLICATION OF IOT**

IoT applications span a wide range of industries, including smart cities, agriculture, manufacturing, health, and emergency response, among many others.

#### **SMART CITIES :-**

In order to make cities smarter and improve general infrastructure, the Internet of Things is essential. When building smart cities, some IoT application areas include intelligent building practices, traffic congestion waste management, intelligent lighting, intelligent parking, intelligent transportation systems, and urban mapping. Installing sound monitoring equipment in sensitive areas of cities, keeping an eye on the number of cars and pedestrians, monitoring vibrations and the material state of bridges and buildings, and monitoring parking spaces available within the city are just a few examples of the various functions that may fall under this category. Smart cities can monitor, regulate, and lessen traffic congestion by utilizing IoT supported by artificial intelligence (AI). [5]

#### **HEALTH CARE :-**

The majority of healthcare systems throughout numerous nations are inherently slow, inefficient, and prone to mistakes. Given that the healthcare industry depends on a variety of activities and gadgets that technology allows for automation and improvement of, The healthcare industry would be significantly altered by more technology that might support a variety of tasks like record keeping, prescription distribution, and report sharing with numerous people and locations [6]. Tracking the flow of patients can greatly enhance hospital workflow. When IoNT is used in the human body, data from in situ body areas that were previously inaccessible for therapy can now be more easily accessible through the use of medical devices

equipped with large sensors. Therefore, IoNT will make it possible to gather fresh medical data, which will result in discoveries and improved diagnostics.

### **SMART AGRICULTURE AND WATER MANAGEMENT :-**

The Internet of Things (IoT) would make it possible to manage and preserve the amount of vitamins present in agricultural products and to control microclimate conditions to maximize the yield and quality of fruits and vegetables. Additionally, monitoring meteorological data enables the forecasting of ice, drought, wind shifts, rain, or snow, allowing temperature and humidity levels to be regulated to avoid fungus and other microbiological pollutants. With regard to cattle, IoT can help identify animals that graze in open spaces, identify harmful gases from animal waste on farms, and regulate the growth circumstances of the progeny to improve.

### **IOT- KEY FEATURES**

IoT basically turns everything "smart," enhancing life in all its facets through the use of networks, artificial intelligence algorithms, and data collection. This might be as easy as adding sensors to your cupboards and refrigerator to recognize when milk and your favorite cereal are running low and to place an order with your favorite grocery store accordingly.

**Interconnectivity:**

Networks are no longer only dependent on large suppliers, thanks to new enabling technologies for networking, and particularly for Internet of Things networking. Networks are still useful even at much lower and less expensive scales. These tiny networks are formed by IoT between its system devices.

**Sensors:**

Without sensors, IoT becomes indistinguishable. They serve as defining tools that turn the Internet of Things from a typical passive network of devices into an active system that can be integrated into the real world.

**Engaging Actively:**

A large portion of modern communication is linked to technology.

### **SECURITY CHALLENGES IN IOT**

The defense against attacks on Internet of Things devices is known as IoT security. IoT device security is less of a known concern, and protection from it is all too frequently disregarded, despite the fact that most business owners are aware that they must use antivirus software on their PCs and phones. Devices using the Internet of Things are widely used. A growing number of items in our environment are being linked to the internet, ranging from refrigerators and cars to surveillance equipment on production lines. The rate of growth of the Internet of Things sector is astounding.

**1. Accuracy of data:**

Through the Internet of Things, billions of gadgets are part of a global ecosystem. Every piece of data that is communicated and transferred between the sensor and the main server can be manipulated, starting with just one data point. Implementing digital signatures and a decentralized distributed ledger is necessary to guarantee integrity [7].

**2. Capacity to Encrypt:**

It takes constant effort to encrypt and decrypt data. In order to process data, IoT network sensors are still lacking. Putting devices on separate networks and installing firewalls will stop brute-force attacks.

**3. Concerns about Privacy:**

Data sharing between platforms, devices, and users is at the heart of the Internet of Things. Data is collected by smart gadgets for several

**4. Integration:**

The number of IoT devices that businesses must manage will eventually increase. Managing such a massive volume of consumer data might present challenges. Undoubtedly, a solitary error or algorithmic infringement is sufficient to bring down the entire data infrastructure [8].

5. Revised:

Compliance is required for managing millions of devices' updates. Updates must be performed manually on devices because not all of them offer over-the-air updates. All of the different gadgets will require one to stay on top of the updates that are available. This becomes a laborious and intricate procedure, and any errors made during the process will result in weaknesses.

## **ETHICS**

This Morals research report was written entirely on its own and wasn't accepted for publication in a journal or conference.

## **SUMMARY**

The increasing number of Internet users and the development of the Internet of Things have led to a demand for data compression across the Internet as a critical function of data proliferation from sensors. Among these are textual data. A brief and clear summary can be produced from a single text or from a collection of texts using summarization, which is a useful method for data aggregation in natural language data processing. Multi-document summarization research is becoming more important in the IoT era because of the scattered location of documents [9]. Although the integration of IoT and summarization approaches has several benefits, this field is still relatively new and has few published studies. [10] Traditional methodologies contain a wealth of prior knowledge that would aid in the model optimization process, and deep neural models have significant non-linear mapping skills.

## **CONCLUSION**

At every stage, the IoT framework is susceptible to attacks. As a result, there are several security risks and demands that must be addressed. The present status of IoT research is mostly focused on access control and authentication protocols; however, in order to achieve the progressive mash-up of IoT topology, new networking protocols like IPv6 and 5G must be consolidated due to the rapid advancement of technology. This chapter's primary goal was to draw attention to the significant security risks associated with IoT, with a special emphasis on security attacks and their defenses. Many IoT devices become soft targets as a result of security flaws, and even in these cases, the victim is unaware that their device is compromised.

## **REFERENCE**

- 1) Sarfraz Alam, Mohammad M. R. Chowdhury, and Josef Noll (2010) SenaaS: An Event-Driven Sensor Virtualization Approach for the Internet of Things.
- 2) Daoliang Li, Yingyi Chen, Oct. 2010, Computer and Computing Technologies Huansheng Ning, Hong Liu, 2012, Cyber-Physical-Social-Based Security Architecture for Future Internet of Things, *Advances in the Internet of Things*
- 3) R. Neisse, G. Steri, and G. Baldini, "Enforcement of security policy rules for the internet of things," in *Int'l Conference on Wireless and Mobile Computing, Networking, and Communications (WiMob)*, 165–172, 2014.
- 4) Mirza Abdur Razzaq and Muhammad Ali Qureshi, "Security Issues in the Internet of Things (IoT): A Comprehensive Study," *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 6, 2017.
- 5) M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's intranet of things to a future internet of things: a wireless-and mobility-related view," *Wireless Communications, IEEE*, vol. 17, no. 6, pp. 44–51, 2010.
- 6) C. Hong Song, F. Zhongchuan, and Z. Dongyan, "Security and trust research in m2m systems," in *Vehicular Electronics and Safety (ICVES)*, 2011 IEEE International Conference on. IEEE, 2011, pp. 286–290.
- 7) Cha, Y. Shah, A. U. Schmidt, A. Leicher, and M. V. Meyerstein, "Trust in m2 communication," *Vehicular Technology Magazine, IEEE*, vol. 4, no. 3, pp. 69–75, 2009.
- 8) C. Ma, W. E. Zhang, M. Guo, H. Wang, and Q. Z. Sheng, "Multi-document Summarization via Deep Learning Techniques
- 9) Y.-K. Ji, Y.-I. Kim, and S. Park, "Big Data Summarization Using Semantic Future for IoT on the Cloud," *Contemporary Engineering Sciences*
- 10) S. Nithyakalyani and S. S. Kumar, "Data Aggregation in Wireless Sensor Networks Using Node Clustering Algorithms: A Comparative Study," in *Proc. of the 2013 IEEE Conference on Information and Communication Technologies (ICT 2013)*, Bandung, Indonesia, 2013.

## Introduction to Artificial Intelligence & Its Applications

**Mr. V. N. Mohod, Mr. A. S. Deshmukh**

Department of Computer Application (BCA)  
vnmohod@gmail.com, anujdeshmukh816@gmail.com  
Vidya Bharati Mahavidyalaya, Amravati.

### ABSTRACT:

Artificial Intelligence (AI) is like the brainpower behind smart machines and computer programs. It's about making computers smart, even if it doesn't exactly mimic human thinking. AI involves using computers to understand and do smart things. Although there's no one-size-fits-all definition, we can think of AI as the study of making computers perceive, reason, and take action. Nowadays, we create a ton of data, more than we can handle. AI is crucial for helping us make sense of all that information and make smart decisions. This paper explores the basics of AI, covering what it is, a bit of its history, how it's used, and its growth and achievements."

### INTRODUCTION:

Artificial Intelligence, often abbreviated as AI, is the field of computer science dedicated to creating intelligent machines and computer programs. These machines are designed to perform tasks that typically require human intelligence, such as learning, problem-solving, understanding natural language, and making decisions. In simpler terms, AI is like giving computers the ability to think and act smartly.

The goal of AI is to develop systems that can mimic or replicate human intelligence, although AI doesn't have to strictly follow how humans think or learn. It involves creating algorithms and models that enable machines to analyze data, recognize patterns, and make decisions, essentially allowing them to perform tasks without explicit programming for each step.

While AI has been around for decades, recent advancements in computing power and data availability have propelled it to the forefront of technology. AI is now an integral part of various applications, from virtual assistants like Siri and Alexa to advanced systems in industries such as healthcare, finance, and transportation. As we continue to explore and expand the capabilities of AI, it is becoming an increasingly important and transformative force in our technological landscape.[1]

### ARTIFICIAL INTELLIGENCE METHODS:

Artificial Intelligence (AI) encompasses a variety of methods and techniques aimed at enabling machines to perform tasks that typically require human intelligence. Here are some key AI methods with brief descriptions:[5]

#### Machine Learning (ML):

Description: Machine Learning involves training a computer system to learn from data and improve its performance on a specific task without being explicitly programmed.

Application: Used in various fields such as image and speech recognition, recommendation systems, and predictive analytics.

**Deep Learning:**

Description: A subset of machine learning that uses artificial neural networks to model and process complex patterns in data. Deep learning has proven particularly effective in handling large datasets.

Application: Commonly used in image and speech recognition, natural language processing, and autonomous systems.

**Natural Language Processing (NLP):**

Description: NLP focuses on enabling computers to understand, interpret, and generate human language. It involves tasks like language translation, sentiment analysis, and text summarization.

Application: Chatbots, language translation services, and voice-activated assistants like Siri or Alexa.

**Computer Vision:**

Description: Computer Vision enables machines to interpret and make decisions based on visual data. It involves tasks such as image recognition, object detection, and facial recognition.

Application: Autonomous vehicles, surveillance systems, and medical image analysis.

**Reinforcement Learning:**

Description: Reinforcement Learning is a type of machine learning where an agent learns by interacting with its environment. The agent receives feedback in the form of rewards or penalties, helping it improve its decision-making over time.

Application: Game playing (e.g., AlphaGo), robotic control systems, and optimization problems.

**Expert Systems:**

Description: Expert Systems are computer programs designed to emulate the decision-making ability of a human expert in a specific domain. They use predefined rules and knowledge bases to provide recommendations or solutions.

Application: Medical diagnosis, troubleshooting technical issues, and decision support systems.

**Genetic Algorithms:**

Description: Genetic Algorithms are optimization algorithms inspired by the process of natural selection. They evolve and refine solutions to find optimal or near-optimal outcomes.

Application: Optimization problems in engineering, finance, and logistics.

**Knowledge Representation and Reasoning:**

Description: Involves representing knowledge in a form that a computer can utilize to solve complex problems. Reasoning mechanisms allow systems to draw conclusions from the represented knowledge.

Application: Expert systems, semantic web, and knowledge-based decision support.

These methods often overlap and can be combined to create more robust AI systems capable of addressing a wide range of challenges in different domains.

### **APPLICATIONS of AI:**

Artificial Intelligence (AI) has a diverse range of applications across various industries, transforming the way we live and work. Here are some notable applications of AI:[4]

#### **Healthcare:**

**Diagnostic Assistance:** AI systems assist in medical diagnosis by analyzing medical images, detecting patterns, and identifying potential health issues.

**Drug Discovery:** AI accelerates the drug discovery process by analyzing large datasets to identify potential compounds and their effects.

#### **Finance:**

**Algorithmic Trading:** AI algorithms analyze market trends and execute trades at high speeds, making financial decisions based on real-time data.

**Fraud Detection:** AI systems detect unusual patterns in financial transactions to identify and prevent fraudulent activities.

#### **Retail:**

**Recommendation Systems:** AI-powered recommendation engines analyze customer preferences and behaviors to suggest products or services, enhancing the overall shopping experience.

**Inventory Management:** AI optimizes inventory levels by predicting demand patterns, reducing waste, and improving supply chain efficiency.

#### **Autonomous Vehicles:**

**Self-Driving Cars:** AI enables vehicles to perceive their surroundings, make real-time decisions, and navigate safely without human intervention.

**Traffic Management:** AI is used to optimize traffic flow, reduce congestion, and enhance overall transportation efficiency.

#### **Education:**

**Personalized Learning:** AI customizes educational content based on individual student needs, adapting to different learning styles and paces.

**Automated Grading:** AI systems automate grading and assessment tasks, saving time for educators.

#### **Customer Service:**

**Chatbots and Virtual Assistants:** AI-powered chatbots provide instant customer support by understanding and responding to user queries.

**Sentiment Analysis:** AI analyzes customer feedback and sentiments to gauge satisfaction levels and improve service quality.

**Manufacturing:**

**Predictive Maintenance:** AI analyzes data from sensors and equipment to predict when machinery is likely to fail, enabling proactive maintenance and minimizing downtime.

**Quality Control:** AI systems inspect and detect defects in real-time during the manufacturing process.

**Cybersecurity:**

**Threat Detection:** AI identifies and responds to cybersecurity threats by analyzing patterns, anomalies, and suspicious activities in network data.

**Fraud Prevention:** AI algorithms help prevent online fraud by identifying unusual transaction patterns and flagging potentially fraudulent activities.

**Human Resources:**

**Recruitment Automation:** AI streamlines the hiring process by analyzing resumes, conducting initial screenings, and identifying suitable candidates.

**Employee Engagement:** AI tools can gauge employee satisfaction and provide insights to improve workplace environments.

**Agriculture:**

**Precision Farming:** AI analyzes data from sensors, satellites, and drones to optimize crop management, irrigation, and harvesting processes.

**Crop Monitoring:** AI helps monitor and diagnose plant diseases, ensuring early intervention for healthier crops.

These applications highlight the versatility of AI, demonstrating its potential to enhance efficiency, accuracy, and decision-making across numerous domains.

**FUTURE of AI:**

The future of Artificial Intelligence (AI) is poised to bring about significant transformations across various sectors, shaping the way we live and work. Here are key trends and possibilities expected in the future of AI:[6]

**AI in Healthcare Revolution:**

AI will play a pivotal role in personalized medicine, drug discovery, and diagnostics. Expect advancements in predictive analytics and AI-powered tools for early disease detection and treatment.

**Autonomous Systems and Robotics Growth:**

Continued development of autonomous systems and robots will impact industries like transportation, manufacturing, and logistics. Self-driving cars, drones, and intelligent robots are likely to become more prevalent.

**AI-driven Personalization:**

AI will further enhance personalized experiences across industries, from e-commerce recommendations and content streaming to personalized education and healthcare plans.



**Quantum Computing and AI Synergy:**

The integration of quantum computing with AI is anticipated to revolutionize the speed and efficiency of complex computations, solving problems that were previously infeasible with classical computing.

**Explainable AI (XAI):**

The need for transparency in AI decision-making will drive the development of Explainable AI, ensuring that AI systems provide understandable explanations for their outputs.

**AI in Edge Computing:**

Edge AI will become more prevalent, enabling processing to occur on local devices rather than relying solely on centralized cloud servers. This could lead to faster response times and increased privacy.

**Ethics, Regulation, and Responsible AI:**

As AI becomes more widespread, there will be a growing focus on ethical considerations, regulations, and responsible AI practices to ensure fairness, accountability, and transparency in AI applications.

**AI-assisted Creativity:**

AI will increasingly collaborate with humans in creative endeavors, assisting in art, design, music composition, and content creation. AI tools will become creative partners rather than just automation tools.

**Natural Language Processing (NLP) Advancements:**

Improvements in NLP will result in more sophisticated language understanding and generation capabilities, leading to more natural and context-aware interactions with AI systems.

**AI for Sustainable Development:**

AI will be utilized to address global challenges, such as climate change, resource management, and sustainable development. AI applications can optimize energy consumption, improve agriculture practices, and enhance environmental monitoring.

**Human-AI Collaboration:**

Collaboration between humans and AI will deepen, emphasizing the augmentation of human capabilities rather than replacement. AI systems will work alongside humans, enhancing productivity and decision-making.

**AI in Education Evolution:**

AI will continue to revolutionize education through personalized learning experiences, intelligent tutoring systems, and adaptive curriculum design to cater to diverse learning styles.

**Enhanced Cybersecurity with AI:**

AI will play a crucial role in cybersecurity by improving threat detection, response times, and developing adaptive defenses against evolving cyber threats.

While these trends suggest a promising future for AI, it's important to navigate challenges related to ethics, bias, security, and societal impacts. Striking a balance between innovation

and responsible deployment will be crucial for ensuring the positive and sustainable growth of AI technologies.

**Conclusion:**

We have covered the basics of Artificial Intelligence, its principles, applications, and achievements. The big aim for those working on AI is to solve tough problems that humans find tricky. The progress in this computer science field is set to change how the world works. Now, it's up to skilled engineers to keep pushing this field forward.

**References:**

1. [http://en.wikibooks.org/wiki/Computer\\_Science:Artificial\\_Intelligence](http://en.wikibooks.org/wiki/Computer_Science:Artificial_Intelligence)
2. <http://www.google.co.in>
3. <http://www.library.thinkquest.org>
4. <https://www.javatpoint.com/application-of-ai>
5. <https://www.educba.com/artificial-intelligence-techniques/>
6. <https://builtin.com/artificial-intelligence/artificial-intelligence-future>

## Study of Wireless Communication Technologies with IoT

**Ms. Vaishnavi T. Chore**

vaishnavichore89@gmail.com

Department of Computer Application (BCA)  
Vidya Bharati Mahavidyalaya, Amravati.

**Prof. V.N.Mohod**

vnmohod@gmail.com

Department of Computer Application(BCA)  
Vidya Bharati Mahavidyalaya, Amravati.

### Abstract:

IOT of Things has gained the attention of almost everybody due to its capability of monitoring and controlling the environment. IoT helps making decisions supported by real data collected using large number of ordinary day-to-day devices that have been augmented with intelligence through the installation of sensing, processing and communication capabilities. One of the main and important aspects of any IoT device is its communication capability for transferring and sharing data between other devices. IoT devices mainly use wireless communication for communicating with other devices. The industry and the research community have proposed many communication technologies for IoT Systems.

### Introduction:

The Internet of Things (IoT) starts with connectivity, but since IoT is a widely diverse and multifaceted realm, you certainly cannot find a one-size-fits-all communication solution. In the past decades, factory automation has been developed worldwide into a very attractive research area. It incorporates different modern disciplines including communication, information, computer, control, sensor and actuator engineering in an integrated way, leading to new solution, better performance and complete system. One of the increasingly important components in factory automation is the industrial communication. For interconnection purposes, a factory automation system can be combining with various sensors, controllers and heterogeneous machines using a common message specification. Many different network type have been promoted for use on an including control area network (CAN)[7]. On the other hand, for accessing networks and services without cables, wireless communication is a fast-growing technology to provide the flexibility and mobility. Reducing the cable restriction is one of the benefits of wireless with respect to cabled devices other benefits include The dynamic network formation, low cost and easy deployment. General speaking, the short range wireless scene is currently held by four protocols, The Bluetooth, UWB (Ultra Wide Band), ZigBee, and Wi-Fi which are corresponding to the standards layer for wireless communication over and action range around 10 to 100 meters.

**Bluetooth:** Bluetooth, standardized by the Institute of Electrical and Electronics Engineers, is originally created by Nokia during the late 90's as an in-house project. However, it quickly became a popular wireless technology that is primarily used for communications between portable devices distributed in a small area (a maximum of 100m coverage range)[11]. Technically, Bluetooth sends short data packets over several channels of bandwidth 1MHz between 2.402GHz to 2.480GHz and its data rate varies from 1Mbps to 3Mbps [11]. Unlike classic Bluetooth optimized for continuous data streaming, BLE is optimized for short data transmissions. In Bluetooth 5.0, enhancements upon BLE's data rates and range were presented by using increased transmit power or coded physical layer. Compared to Bluetooth 4.0, maximum 4x transmission range increase is expected and a maximum data rate of 2Mbps can be achieved (as twice as fast) [11]. In the latest Bluetooth, direction finding feature of BLE was

enhanced to better understand signal direction and achieve sub-meter location accuracy [11]. To enable large-scale IoT device networks that support many-to-many device communications, BLE mesh networking has been adopted in 2017. A relay device only relays a message if the message is not in the cache and its TTL is greater than 1 [11]. Each time message is received and retransmitted; TTL will be decremented by one. If the TTL reaches zero, the message will be discarded at the relay device, eliminating endless loops. In addition, the backwards compatibility feature and friendship feature are also defined in BLE mesh for BLE devices.

**Application of Bluetooth:** The transfer of files like images, mp3 audio is very easy in the mobile phones. A little amount of bandwidth is needed for the wireless networking between the laptops and desktop computers. All the peripheral devices are like mouse, keyboard, printers, speakers, etc. are connected to the PC cordlessly.

**Advantages of Bluetooth:** It has low power consumption and it can pass through the walls. The Bluetooth has Range better than infrared communication. The technology is adopted in many products such as head set, in car system, printer, web cam, GPS, system, keyboard and mouse. Due to availability of Bluetooth head phones, call can be taken on phone even while driving and doing some other activity simultaneously. This Hand Free operation relieves great strain and It is use for voice and data transfer.

**Disadvantages of Bluetooth:** It has low bandwidth as compared to Wi-Fi. The Battery usage is more compare to the condition when Bluetooth is powered OFF. The new technology known as BLE or bluetooth low energy or Bluetooth smart is developed to enhance the battery life further.

One of the big disadvantages of Bluetooth is security. This is due to the fact that it operates on Radio frequency and hence can penetrate through walls. It is advisable not to use it for critical business or personal data transfer.

**ZigBee:** ZigBee is another short-range wireless technology for wireless personal area networks (WPAN), ZigBee has been widely considered for a variety of IoT applications including home automation, industrial monitoring, and health and aging population care. Similar to BLE, ZigBee is also a low power technology. ZigBee operates in the unlicensed bands, i.e., mainly at 2.4GHz and optionally at 868MHz or 915MHz, and its default operation mode at 2.4GHz uses 16 channels of 2MHz bandwidth. ZigBee is able to connect up to 255 devices at a time with a maximum packet size of 128bytes. Depending on the blockage of environments, the transmission ranges between devices vary from a few meters up to 100 meters.[9]The end devices are logically connected to a coordinator or routers. However, these end devices cannot directly communicate with each other. To enable large-scale IoT device networks, ZigBee can be extended as generic mesh where devices are clustered with a local coordinator and connected via multihop to a global coordinator. In addition, while BLE allows four different data rates varying from 125kbps to 2Mbps, ZigBee can only support data rates from 20kbps to 250kbps. According to the performance evaluation in a realistic home automation scenario in[6].

**How Zigbee works?** Zigbee system structure consists of three different types of devices such as Zigbee coordinator, Router, and End device. Every Zigbee network must comprise of at least one coordinator who acts as a root and bridge of the network. The number of routers, coordinators, and end devices depends on the type of networks such as star, tree, and mesh networks. A network coordinator is a device that sets up the network, is aware of all the nodes within its network, and manages both the information about each node as well as the information that is being transmitted/ received within the network. Every ZigBee network must contain a network coordinator [9]. Mesh networking allows for redundancy in node links so that if one node goes down, devices can find an alternative path to communicate with one another.

**Application of ZigBee:**

**Smart Metering:** Zigbee remote operations in smart metering include energy consumption response, pricing support, security over power theft, etc [6].

**Home Automation:** Zigbee is perfectly suited for controlling home appliances remotely as a lighting system control, appliance control, heating, and cooling system control, safety equipment operations and control, surveillance. [9].

**Industrial Automation:** In manufacturing and production industries, a communication link continually monitors various parameters and critical equipment. Hence Zigbee considerably reduces this communication cost as well as optimizes the control process for greater reliability[9].

**Smart Grid monitoring:** Zigbee operations in this smart grid involve remote temperature monitoring, fault locating, reactive power management, and so on.

**Advantages of ZigBee:** It will take the place of existing infrared technology based devices. This will save cost of battery replacement as ZigBee uses lithium battery which last long [9]. It is easy to monitor and control home appliance from remote. Their Setting up the network is very simple, easy and It does not have central controller and load are distributed evenly across the network [6]. The network is scalable and it is easy to add / remove ZigBee end device to the network.

**Disadvantages of ZigBee:** As compared with Wi-Fi, it is not secure. The high replacement cost once any issue happens within Zigbee based home appliances. The transmission rate of the ZigBee is less [6]. It does not include several end devices. It is so highly risky to be used for official private information. It is not used as an outdoor wireless communication system because it has less coverage limit [9].

**Future scope of ZigBee technology:** Zigbee has very promising future in front of it. Its various types of areas such as defense, national security, monitoring and controlled etc. can be facilitated by device based on zigbee standard [9]. It leads to the cheap wireless technology, so that it can be widely use for low rate data transfer. Zigbee aims to achieve greater efficiency. By rapid rise in home networking, zigbee world provide revolutionizing statistics in the upcoming year which would entirely change to the wireless world.

**Wi-Fi:** Wi-Fi is a family of technologies commonly used for wireless local area networks (WLAN). Different from Bluetooth and ZigBee that provide connectivity between devices, Wi-Fi provides the last mile wireless broadband connections for devices to the Internet with a larger coverage and higher data rates [8]. In fact, Wi-Fi has been evolved several generations to support higher throughputs. Specifically were introduced in 1999, where it can support a data rate up to 54Mbps in 5GHz, and was released with a maximum data rate of 54Mbps in 2.4GHz. standards were not able to meet the growing demand of hypermedia applications over WLANs due to their relatively low throughputs and capacity. [5] higher data rates (up to 600Mbps and 7Gbps) with a wider coverage compared to previous ones by using dense modulations and MIMO technology. In addition,(Wi-Fi HaLow) was introduced in 2017to support IoT with extended. coverage and low-power consumption requirements. It operates in the unlicensed sub-1GHz bands(excluding the TV white-space bands) and its bandwidth occupation is usually only 1MHz or 2MHz, while in some countries, wider bandwidths up to 16MHz are also allowed. Compared to high-speed Wi-Fi generations, the aims to provide connectivity to thousands of devices with coverage of up to 1km but its maximum data rate is about 300Mbps utilizing 16MHz bandwidth [4], [7],[5],[8].

**Application of Wi-Fi:** Send documents to your printer from any computer or Smartphone. Stream movies to any TV in the house. Forward notifications from your Smartphone to your

---

Pc. Share file with nearby computers. Turn your Smartphone into a remote control [7]. Sync your music library, photo library, or other files with your Smartphone USB-free.

**Advantages of Wi-Fi:** Wireless networks allow users to access network resources from nearly any convenient location within their primary networking environment (a home or office). The increasing saturation of laptop-style computers, this is particularly relevant. With the emergence of public wireless networks, users can access the internet even outside their normal work environment. Wired networks, on the other hand, have the additional cost and complexity of actual physical cables being run to numerous locations (which can even be impossible for hard-to-reach locations within a building) [5]. Wireless networks can serve a suddenly-increased number of clients with the existing equipment. In a wired network, additional clients would require additional wiring. Wireless networking hardware is at worst a modest increase from wired counterparts. This potentially increased cost is almost always more than outweighed by the savings in cost and labor associated to running physical cables [8].

**Disadvantages of Wi-Fi:** To combat this consideration, wireless networks may choose to utilize some of the various encryption technologies available. Some of the more commonly utilized encryption methods, however, are known to have weaknesses that a dedicated adversary can compromise [4]. The typical range of a common 802.11g network with standard equipment is on the order of tens of meters. While sufficient for a typical home, it will be insufficient in a larger structure. To obtain additional range, repeaters or additional access points will have to be purchased. Costs for these items can add up quickly. Any radio frequency transmission, wireless networking signals are subject to a wide variety of interference, as well as complex propagation effects that are beyond the control of the network administrator. The speed on most wireless networks (typically 1-54 Mbps) is far slower than even the slowest common wired networks (100Mbps up to several Gbps). However, in specialized environments, the throughput of a wired network might be necessary [7].

**5G Technology:** 5G Technology stands for 5th generation mobile technology. 5G represent the next major phase of mobile telecommunication ethics beyond the upcoming 4G standards. 5G technology is contribution the service in Product Manufacturing, Documentation, supporting electronic communications, etc [4].

**Advantages of 5G Technology:** In 5G Tech. High determination and bi-directional large bandwidth shaping. This Technology to wrinkle all networks on one platform. Its an More active and effective. Technology to simplify subscriber administration tools for the quick action.

**Disadvantages of 5G Technology:** Many of the old devices would not be able to 5G, hence, all of them need to be swapped with a new one expensive deal. Developing infrastructure needs high cost. Security and privacy problems yet to be solved.

**Conclusion:**

Wireless communication technology has played a great role in the energy Internet and improved the intelligent level of energy network control. The major functions of wireless communication equipment include convenience, positioning, remote communication and control. Wireless communication can improve the efficiency of power grid operation and maintenance. with the wireless communication system for energy, the transmission lines and distribution facilities can be monitored through ultra-high definition cameras, so as to find hidden faults in time and save manpower and material resources for on-site inspection. The characteristics of advanced wireless communication systems with large bandwidth, low delay, wide connection, and high reliability fully meet the communication needs of an energy Internet. One can make full use of wireless communication technology to develop new solutions, which can meet customer needs successfully in energy networks with high penetration.

---

**References:**

- 1) C. Fang, K. Dou, J. Liu, W. Tao and M. Ma, "Analysis of Synchronous Pharos Data Wireless Communication Technology for Distribution Network," *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing, China, 2018, pp. 1-5,doi: 10.1109/EI2.2018.8582330.
- 2) International Journal for Research Trends and Innovation (www.ijrti.org)
- 3) "Badge" that enables staff communication. *Br J Healthcare Compute Inform Manage* 2006 (May);23(4). <http://www.bjhcim.co.uk/news/industry/2006/ind605003.htm>. Accessed February 11, 2008
- 4) Q. Wu, "4G Communication Technology Wireless Network Secure Communication," *2021 International Wireless Communications and Mobile Computing (IWCMC)*, Harbin City, China, 2021, pp. 915-918, doi: 10.1109/IWCMC51323.2021.9498797.
- 5) L. Ma *et al.*, "Application of Wireless Communication Technology in Ubiquitous Power Internet of Things," *2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET)*, Beijing, China, 2020, pp. 267-271, doi: 10.1109/CCET50901.2020.92131
- 6) W. Wang, G. He and J. Wan, "Research on Zigbee wireless communication technology," *2011 International Conference on Electrical and Control Engineering*, Yichang, China, 2011, pp. 1245-1249, doi: 10.1109/ICECENG.2011.6057961.
- 7) C. Li, Y. Tang and F. Luo, "Research on Internet Regional Security Based on 5G Wireless Communication Technology," *2022 International Conference on Smart City and Green Energy (ICSCGE)*, Hong Kong, Hong Kong, 2022, pp. 1-4, doi: 10.1109/ICSCGE58152.2022.00008.
- 8) Q. Zhang, L. He, H. Yang, Y. Li, X. Duan and W. Jiang, "5G key technology and deep application in power system," *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China, 2022, pp. 863-867,doi: 10.1109/ITAIC54216.2022.9836335.
- 9) S. Long and F. Miao, "Research on ZigBee wireless communication technology and its application," *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chengdu, China, 2019, pp. 1830-1834,doi: 10.1109/IAEAC47372.2019.8997928.
- 10)S. Pradhan, E. Lawrence and A. Zmijewska, "Bluetooth as an enabling technology in mobile transactions," *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, Las Vegas, NV, USA, 2005, pp. 53-58 Vol. 2,doi: 10.1109/ITCC.2005.98

## Role of Nanotechnology and Artificial Intelligence in aroma

**Ms. Mayuri A. Deshmukh**

Asst. Professor, Department of Computer Science,  
Vidya Bharati Mahavidyalaya, Amravati

### **Abstract:**

Artificial intelligence is most occupied concept in world of technology, its presence in almost any industry that deals with any huge volume of data are taking advantages in day to day operation. This technology can be applied to many different sectors and industries. There has been a tremendous use of artificial intelligence in nanotechnology research during last decades. Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. The application of artificial intelligence and nanotechnology in cosmetics has been shown to overcome the drawbacks associated with traditional cosmetic and also to add tremendous development in this field. Aromatherapy is both an Art and a Science. Aromatherapy is generally used for treatment of common problems like headache, cough, cold, sleep related issues, stress, anxiety etc. Advanced practitioners are using aromatherapy in treating psychological as well as various physiological issues. The AI technology can be used to analyse the current data inferences, patterns and learning and use them to create new blends of flavours and fragrances which can then be used. Hence the industry and technology is leading us India should aim at developing its own technology which will help several small perfumery companies to utilize the new program in development of the new trends in fragrances. The aim of this paper is to create development in field of aroma with new technologies.

### **Keywords:**

Artificial Intelligence (AI), Machine learning, Nano technology, aroma

### **Introduction:**

Lifestyle is a way used by people, groups and nations and is formed in specific geographical, economic, political, cultural and religious text. Lifestyle is referred to the characteristics of inhabitants of a region in special time and place. In recent decades, life style as an important factor of health is more interested by researchers. According to WHO, 60% of related factors to individual health and quality of life are correlated to life style [1]. Today, wide changes have occurred in life of all people. Malnutrition, unhealthy diet, smoking, alcohol consuming, drug abuse, stress and so on, are the presentations of unhealthy life style that they are used as dominant form of lifestyle. Besides, the lives of citizens face with new challenges. Therefore, according to the existing studies, lifestyle has a significant influence on physical and mental health of human being.

Fragrance is a sweet or pleasant odor or scent, there are some natural ingredients having a pleasant fragrance. Nanotechnology has entered the production and application of various personal care and cosmetics products. Nano perfume ejectors are designed to mix nanoparticles with perfume and/or water particles and enable sterilization of air, absorption of unpleasant, and release of pleasant odors. Electronic noses are used to study the use of nanoparticles in fragrant products.

These days nanotechnology is small things such as an atom the "Size of the Nano scale" or basically, just how small is "nano?" and what can we imagine about the scale of from



Microscopic perspective. From metric MKS unit dimensional point of view or International System Units (ISU), the prefix “nano” means one-billionth or  $10^{-9}$ ; therefore, one nanometre is one-billionth of a meter. It’s difficult to imagine just how small that is, thus, here we are presenting some examples to clear the matter better [2].

1. A sheet of paper is about 100,000 nanometres thick.
2. A human hair is approximately 80,000- 100,000 nanometres wide.
3. A single gold atom is about a third of a nanometer in diameter.
4. On a comparative scale, if a marble diameter were one nanometre, then the diameter of the Earth would be about one meter.
5. One nanometre is about as long as your fingernail grows in one second.
6. A strand of human DNA is 2.5 nanometres in diameter.
7. There are 25,400,000 nanometres in one inch.

Nano-science and nanotechnology involve the ability to see and to control individual atoms and molecules. Nanotechnology is the use of matter on an atomic, molecular, and supra molecular scale for industrial purposes, and it is the design, production, and application of structures, devices, and systems by manipulation of size and shape at the nanometre scale. Thus, when it comes to Nano science and nanotechnology, we can say that when and where, we are dealing with smallest scale size, “Small is Powerful”. However, when we are dealing with small, in particular at the scale of atom size, then for the purpose of all practical situations and applications, in particular in field medicine, we are encountering, with share volume of data that we need to collect and be able to analyse these data to the point of real-time speed. So requires a lot of data analytics and data mining, where such practice is beyond human capacities.

Artificial intelligence has been an increasingly growing area for many decades now, not just within itself where the areas of Machine learning, Deep learning and artificial neural network work simultaneously, but also in the number of fields and industries that they are now prevalent in. Nano-science and nanotechnology are the study and application of tiny things. There are some growing areas where AI converges with nanotechnology. [3] Perfumes are defined as products which give good odour to the product and person on/in which they have been used. But with increasing consumer expectations and awareness the demands from the perfumes have increased. The modern consumer demands are:

1. Multifunctional properties
2. Long lasting results
3. Convenient to use
4. Affordable price

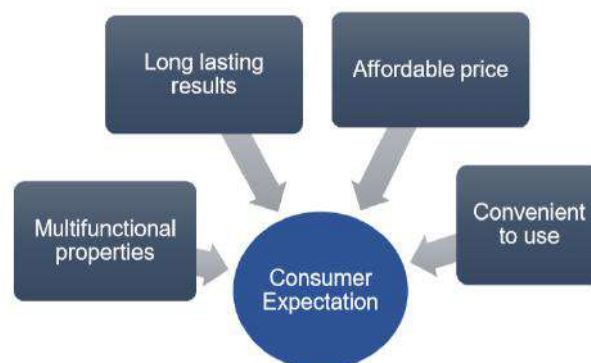


Fig:1 The modern consumer demands

With all these demands the expectations from a perfume company has increased 10 folds over the years. But these expectations have brought with them a challenge for creating an innovative product with best quality, maximum activity and best retentively at very low cost with high end popular expensive raw materials of best quality. Though these aspects look very lucrative from the marketing point of view but from a perfumer's view point they are nightmare as one must fit the best gems in a diamond encrusted platinum box at the price of a cardboard box. To maintain this balance a perfumer generally looks at various aspects like cheap raw materials or haggling on costs of ingredients or using less concentration of other expensive raw materials but forgets to look at increasing the availability of the expensive ingredients which helps in increasing the activity and also gives cost benefits. The potency and retentively of the perfumes can be increased by the modern technologies reason behind this is:

1. Achieve targeted & timed release of perfumes
2. Enhance efficacy in terms of retention and strength
3. Provide effective concentration
4. To preserve the stability of actives
5. Minimize irritation potential
6. Higher retentively of fragrances and clarity of product for aesthetically pleasing formulations.

Generally, in skin and other sectors Delivery Systems are commonly used. A delivery system is the method of delivering active payloads on to the surface, and then having them pass through the lipid barrier and finally reaching the targeted lower layers beneath.

### **Nanotechnology:**

Nanotechnology is regarded as the most imminent technology of 21st century and is contemplated as a big boon in the cosmetic industry. The term nanotechnology is the combination of two words: namely, technology and the Greek numerical "nano" which means dwarf. Thus, nanotechnology is considered as the science and technology used to develop or manipulate the particles in the size range of 1 to 100 nm. Since 1959, nanotechnology has emerged in different fields like engineering, physics, chemistry, biology, and science and it has been virtually 40 years since nanotechnology has intruded into the field of cosmetics, health products, and dermal preparations.

There are several advantages of nanotechnology. Some of them are listed below:

- a) Nanotechnology allows for the controlled release of active substances by controlling the drug release from carriers by several factors including physical or chemical interaction among the components, composition of drug, polymer and additives, ratio, and preparation method.
- b) Nano cosmeceuticals make the fragrances last longer, for example, Allure Perfume and Allure Eau Perfume spray by Chanel.
- c) Nanotechnology based formulations are more effective and increase the lasting potential of fragrances by having very small size of the particles, the surface area is increased which allows the active transport of the active ingredients into the skin.
- d) Nanotechnology based cosmetics have high entrapment efficiency and good sensorial properties and are more stable than the conventional cosmetics. Most of the nanoparticles are suitable for both lipophilic and hydrophilic materials.

---

## Global Industry Overview

The global perfumery market reached a value of US\$ 38.8 Billion in 2018. The global perfume market size was estimated at USD 32.50 billion in 2019 and is expected to reach USD 33.69 billion in 2020. The market value is projected to reach US\$ 48.0 Billion by 2024, at a projected CAGR of 3.6% over period of 2019- 2024. The market growth is attributed to the growing trend of personal grooming, coupled with increasing demand for luxury and exotic fragrances. Moreover, increasing consumer spending on premium and luxury fragrances due to the high income level, along with improving living standards, is driving the global market. In recent years, perfumes have evolved into a significant business in the cosmetics and personal care industry. Product diversification by manufacturers is also expected to expand the customer base. Product innovations based on customer needs are further augmenting the sales in the perfume market. For instance, Lauder's Jo Malone stores offer fragrance consultations so that shoppers can develop a customized product. Key players are also focusing on introducing natural fragrances in the premium category, primarily due to rising concerns over allergies and toxins in synthetic ingredients. Premium perfumes are expected to expand at the fastest CAGR of 3.9% from 2019 to 2025 owing to the growing preference for unique, handcrafted, and exotic fragrances.

Market size value in 2020 USD 33.69 billion  
Revenue forecast in 2025 USD 40.9 billion  
Growth Rate CAGR of 3.9% from 2019 to 2025

Though the predictions were made before COVID, after COVID it was seen that the Fragrances and Perfumes market in the U.S. is estimated at US\$11.8 Billion in the year 2020. The country currently accounts for a 27% share in the global market. China, the world second largest economy, is forecast to reach an estimated market size of US\$11.3 Billion in the year 2027 trailing a CAGR of 6% through 2027. Among the other noteworthy geographic markets are Japan and Canada, each forecast to grow at 0.9% and 2.4% respectively over the 2020-2027 period. Within Europe, Germany is forecast to grow at approximately 1.6% CAGR while Rest of European market will reach US\$11.3 Billion by the year 2027. At present several new technologies are being used in the global industry for the creation of long lasting, effective, retentive fragrances with longer shelf life. These are being done with use of new delivery systems like nanotechnology, micro-encapsulations, microsponges, liposomes etc.

Artificial intelligence and data-driven algorithms are heralding a new era for the fragrance industry, which is being transformed by machine learning.

### 1 Symrise's Philyra

Symrise's Philyra, created in partnership with IBM Research, analyzes thousands of formulas in order to identify patterns and discover innovative fragrance combinations. The system's algorithms accelerate the fragrance creation process by designing formulas that have never been seen before. This includes algorithms that learn and predict raw material substitutes and complements that can be used in a formula, appropriate dosing for a raw material based on usage patterns, 'likability' factor (whether the fragrance will be well received), and novelty of the fragrance when compared to commercially available fragrances. Philyra's data-driven approach also leverages data on fragrance families, historical data, and industry trends.

Philyra uses machine learning to discover whitespaces in the global fragrance market and create new formulas. Symrise's perfumers add the final touch by finetuning the creations, for example, by emphasizing a certain note or improving the long lastingness of the fragrance.

---

Philyra created two millennial fragrances that launched in 2019 for Brazilian personal care company O’Boticario.

## 2. Givaudan’s Carto

Givaudan’s Carto is an AI-powered tool that is designed to reinvent the way perfumers create, with the added benefit of accelerating perfume development. The AI program invites perfumers to imagine and create new fragrance accords using an interactive touch screen (creating their formulas differently from the traditional spreadsheet or olfactive pyramids). The program can cross-reference the fragrance house’s own market research, research and development, consumer data and historical formulas. Carto also includes an instant-sampling robot that accelerates the production of fragrance trials.

Carto enables perfumers to experiment with creative concepts by using the playful AI interface, which is supported by an extensive data library of fragrance formulations.

Carto is being used in Givaudan’s fragrance creative centers in all regions.

## 3. Scent bird’s Confessions of a Rebel

When direct-to-consumer Scent bird launched new gender-fluid sub-brand Confessions of a Rebel, it used AI, consumer data and reviews to create its four initial fragrances. (Confessions of a Rebel defines itself as a ‘next-gen fragrance brand, ready to push boundaries’). Scent bird leveraged over a million data points from its 300,000 subscribers to conceive the directions and fragrances. Instead of using typical fragrance categories such as floral, woody, or citrus, Scent bird asked consumers to select their own descriptors, that included ‘fresh’, ‘clean’ and ‘sexy’.

AI software gives Scent bird immediate access to patterns within its extensive subscriber data base, enabling it to create fragrances by way of user reviews, consumer preferences, and fragrance note preferences.

Confession of a Rebel’s four gender-fluid fragrances, including Get A Room, Love High, About Last Night, and Almost Single, launched in mid-July.

## 4. Sommelier du Perfume

Sommelier du Perfume is an AI-powered fragrance app that helps users find their ideal fragrance. Algorithms analyse responses to a questionnaire, learning about users’ tastes and lifestyles in order to make recommendations from its database of over 30,000 fragrances. After users select perfumes that they want to test in-store among the app’s shortlist, nearby retail stores are identified (consisting of both large beauty retailers and independent boutiques). Fragrance information includes olfactory notes, the perfumer and fragrance’s history, and ingredients, together with a toxicity assessment.

Sommelier du Perfume, as the name implies, educates consumers about the breadth of fragrances in the marketplace, and makes the selection accessible to them within a user-friendly app.

Sommelier du Perfume helps consumers find their next fragrance which they can buy in 8,000 stores in the US.

## 5. Algorithmic Perfumery

Algorithmic Perfumery invites users to create their own personalized fragrance made by artificial intelligence. The system uses AI software and a variety of data (together with a sampling robot) to create a personalized fragrance for each person who interacts with it. Algorithmic Perfumery is refined all the time as new users continue to train the creative sensibilities of the AI system, and it adapts and learns from every exchange. Dutch founder

Frederik Duerinck, who started the Netherlands-based company Scenatronix, has presented Algorithmic Perfumery at film festivals and art exhibitions throughout 2019.

Duerinck's goal is to change how users interact with fragrance, and for every user to have their own unique fragrance. This includes fragrances with under-explored olfactive categories that are not typical of commercially successful fragrances.

Algorithmic Perfumery has been on tour this year throughout Europe and North America.

## 6. Coty's VR Experience

Coty's fragrance-focused, multi-sensorial virtual reality experience, launched in Argentina with retailer Julieraque, is powered by AI. The immersive experience uses touch, smell, sound and sight to help consumers find their perfectly fine fragrance match. Shoppers wear a virtual reality headset and pick up a scented stone which activates a short video. Each stone is tied to a broad fragrance category, such as 'citrus watery', 'floral fruity' or 'oriental spicy', up to six recommendations from eight Coty luxury brands based on their favourite fragrance concepts.

The VR technology can be scaled and adapted to suit a variety of markets and brands. The experience merges physical and digital worlds, and helps users navigate the world of fragrances.

Coty is planning on bringing the VR experience to additional markets, and tailoring the experience to specific brand universe In a nutshell if we see these new tools help to

- a) Predict alternative raw material or substitutes to be used
- b) Human response
- c) The novelty of fragrance as well as the appropriate dosing of raw material, among other components.
- d) It also helps in understanding consumer preferences which in turn helps the company to focus on perfecting the final product rather than spending time searching for new fragrance combinations.

This data has been generated by a team of computer scientists who used a set of algorithms to predict the odour of different molecules based on their chemical structure. They labelled the smell with more than 19 descriptors, including "fish," "garlic," "sweet," or "burnt." They also created a massive database based on pleasantness and intensity of odour. While the immediate use of these programs was not sure, it can now find a way into the fragrance or perfume industry. New technology based on artificial intelligence could accurately predict future taste preferences among specific groups, allowing food and drink companies to get ahead of the next big trend and better target new product launches.

Application of Nanotechnology in perfumery is a very important aspect to be considered seriously to grow in the industry. Currently known applications of nanotechnology in perfume production and application are predominantly based on nano-encapsulation methods.

Application of nanotechnology enables:

- Reduction of costs of perfume compounds manufacturing,
- Manufacturing high quality ingredients,
- Manufacturing complete natural perfume compounds since they are derived from reaction catalyzed by enzymes from natural organisms,
- Compounds of low toxicity like using gold nanoparticles to replace toxic reagents that increases oxidation of aromatic primary alcohols to aldehydes,
- Manufacture of highly sensitive perfumery compounds.

## Application of AI in Fragrance Industry in India

Today the Indian Fragrance market it is still lagging in terms of quality Fragrance creation. Indian fragrance industry is still looked down upon by the global industry as lacking in quality and substance. Especially the small companies dealing with the creation of fragrance blends. This is due to the aspect that Indian business owners in this field are still not well versed in this field and lack the advanced technical knowledge in creating these products. After looking at the global data in the fragrance industry we can say that if we bring in the Artificial intelligence technologies in India then they can be used in the Indian scenario to:

- a) Improve the standard of Fragrance Industry in India
- b) Help in improving the quality of fragrance blends
- c) Help the small and medium scale industries to upgrade themselves
- d) Aid the government in creating a tool for the upliftment of fragrance sector in India.
- e) Provide more employment.

Hence the industry and technology is leading us India should aim at developing its own technology which will help several small perfumery companies to utilize the new program in development of the new trends in fragrances. These systems will help in generating a new breed of entrepreneurs who can start selling their own perfume blends which have been created by AI technology and will also help the shopkeepers. The aid of AI technology Fragrance creation Kannauj and the area around it can be converted into a Tourist destination and Fragrance capital of India like Grasse is for France. Today Kannauj and the area is known for the perfumes but still there are no real fragrances and experiences which people can take back. So, with these technologies the small shop owners selling attars can also get help and can start selling the ethnic fragrances with a modern twist, created with the help of these technologies.

The technology can be divided into two parts:

### 1. E nose:-

Use of Artificial olfaction, i.e., e-nose, plays a critical function in robotics by mimicking the human olfactory organ that can recognize different smells. The e-nose through mimicking the olfactory receptors with the programmed algorithm of the artificial neural network helps in the recognition of the pattern of odors (i.e., their chemical profiles). Advantages of E Nose:

- Allows for monitoring of the shelf life of natural perfumery herbs by sensing the aromatic VOCs due to post-harvesting, respiration, fermentation, and phenolic oxidation.
- Detection of off notes in a perfume blend.
- Detection of perfumery compounds in a perfume blend.
- Detection of impact of factors on the stability of fragrance blends.
- Detection of Quality of Raw materials:
  - o Identification of raw materials used in perfumery.
  - o Detection of adulteration in perfumery raw materials.

### 2. Artificial intelligence Technology:-

The AI technology can be used to analyse the current data inferences, patterns and learnings and use them to create new blends of flavours and fragrances which can then be used by the entrepreneurs, Retailers and other small and medium companies to create new quality flavour and fragrances as per the choice of current consumers. The AI system can be created by using Python.

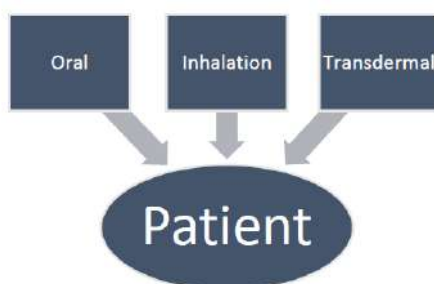
Advantages of AI in Fragrance Industry:

- Creation of new modern blends.
- Will help the entrepreneur in creating world class fragrances.
- Can Create Personalized fragrance and flavours for companies.

- Can help Kannauj to become the fragrance and flavour capital again.
- Can help several retailers to give the consumers exact fragrance and flavour of their choice.
- Cost effective as will help in saving Raw material and energy cost earlier used in trial and errors.
- Will bring in revenue for the government as companies can utilize the capabilities to create unique, personalized, quality fragrance blends.

### **Aromatherapy, AI and Nanotechnology**

Aromatherapy is both an Art and a Science. It has for ages been considered as a holistic form of therapy. Aromatherapy is the use of essential oils for therapeutic or medical purposes (Buckle, 2003). It acts upon the holistic principles of awakening and strengthening energies and promoting self-healing. Buchbauer defined aromatherapy as – “therapeutic uses of fragrances to cure, mitigate or prevent diseases, infections and indispositions by means of inhalation.” The main component of aroma therapy is essential oils. Aromatherapy is generally used for treatment of common problems like headache, cough, cold, sleep related issues, stress, anxiety etc. Advanced practitioners are using aromatherapy in treating psychological as well as various physiological issues. Aromatherapy is generally administered in three ways to a patient:



- a. Oral route: The essential oils are generally administered orally in the form of oils or as seasoning in consumable products. The oils are Bitter tasting and may irritate mucosal lining. This mode of application is suitable only under supervision of experienced practitioners, as regular ingestion over a long period of time may lead to hepatotoxicity. Oils are often formulated into capsules and then consumed orally.
- b. Transdermal route: Lipid solubility of essential oils allows for better penetration in the skin. Easy penetration generally happens maximum behind ears, eyelids, inside the wrist followed by soles, palms, forehead, scalp, armpits, and the least in legs, buttocks, trunk and abdomen. The oils are often applied by rubbing or massaging. The oils are generally incorporated in carrier oils, gels or creams and then massaged on the skin.
- c. Inhalation: This is the most common and effective route of administration and is regarded as “true aromatherapy”. The incidence of adverse effects using this method is very rare.

Methods of administration include spraying on cloth and using an aroma diffuser, among others.

Essential oils: The main component of aroma therapy is essential oils, also sometimes referred to as volatile oils. Essential oils are aromatic oils extracted naturally from plant and animal sources and are used for treatments. Essential oils play a key role in plant metabolism and are also used in plants for communication. They are used by plants to attract certain beneficial insects and repel others. Essential oils allow plants to send and receive signals. Chemical communication requires specific signals that can be clearly recognized and interpreted. These chemicals are mostly in combinations, (like acetal and ester) or in enantiomeric forms e.g.  $\alpha$  bisabolol has two enantiomeric forms - (+)  $\alpha$  bisabolol and (-)  $\alpha$  bisabolol.

Essential oils contain two main groups according to their biochemical origin :

1. Terpenes & Higher Monologues
2. Phenylpropane derivatives (cinnamic acid & aldehyde)

Oils whose main constituents belong to the same group exhibit similar effects, but again each oil exhibits different characteristics apart from these similar effects. Terpenes & Higher homologues (other molecules based on terpenes) - e.g. geraniol (terpene alcohol-antiseptic and tonic), Farnesol (Sesquiterpene alcohol) etc. Essential oils contain mostly mono and sesquiterpenes. Mono terpenes are smaller molecules and their high oil content gives more clarity, less viscosity and more volatility e.g. eucalyptus oil. Sesquiterpenes have larger molecular weight and make the oil coloured (yellow/dark yellow/brown) and more viscous e.g. sandalwood, patchouli. Phenylpropane derivatives (cinnamic acid & aldehyde) - are by-products of the amino acid metabolism. These break down to form substances like anethol (antispasmodic, stabilizing effects), eugenol (stimulant, irritant & antiseptic property) etc.

Aromatherapy mechanism:

Two basic mechanisms are offered to explain the effect of aromatherapy.

1. First is the influence of the aroma in the brain, especially the limbic system through the olfactory system.
2. The other is the direct pharmacological effects of essential oils.

The efficacy of the aromatherapy remains unproven. Once the oils are circulating in the blood, they are carried to the target organ, where they exert a therapeutic effect on the specific tissues. e.g. Juniper oil targets the urinary tract and kidneys, with secondary effects on the Digestive, Respiratory and Reproductive Systems. Chamomile Oil targets the Nervous System via which it exerts a broad effect on body Systems, like Digestive Tract etc.

Examples of therapies applied and absorbed through the skin include stress relieving therapies and motion sickness patches. The global fragrance industry has shown affirmation that essential oils can affect the mood, boost productivity, help in restful sleep, alter psychological conditions of human beings and modify human behaviour on a regular and subconscious level.

Artificial intelligence can be used in aromatherapy for:

1. Evaluating and diagnosing the anxiety levels and problems in patients.
2. Evaluating the chemical constituents of essential oils.
3. Finding out the potential antimicrobial ability of essential oils.
4. Finding out the purity and level of adulteration by E-Nose and further finding out the ability of an oil in treating a certain health condition.
5. AI can be used in creating a shopping app for the customers to help them in selecting the correct oil or combinations of oils for the treatment of particular issue they are facing.

### **Conclusion:**

With the progress and thriving of Artificial Intelligence (AI) and Nanotechnology curves are both on ascending slope, it seems these two curves have narrowed to each other to the point that we can recognize a separation span between them. This indicates that they have two complements each other and be the right partner and companion when it comes to both these industries. In today's technology of artificial intelligence and nanotechnology, it seems their integration of these is inevitable scenario. Aromatherapy is generally used for treatment of common problems like headache, cough, cold, sleep related issues, stress, anxiety etc. Advanced practitioners are using aromatherapy in treating psychological as well as various physiological issues. The AI technology can be used to analyse the current data inferences, patterns and learning and use them to create new blends of flavours and fragrances which can then be used. Hence the industry and technology is leading us India should aim at developing



its own technology which will help several small perfumery companies to utilize the new program in development of the new trends in fragrances.

### Reference:

1. Ziglio E, Currie C, Rasmussen VB. (2004). The WHO cross-national study of health behavior in school aged children from 35 countries: findings from 2001–2002. *J School Health*, 74 (6): 204–206.
2. WHO (2001). Services for prevention and management of genetic disorders and birth defect in developing countries (Farhud DD. As committee member) (WHO/HGN/WAOPB-D/99.1).
3. Farahnaz Behgounia, Bahman Zohuri\*. Artificial Intelligence Integration with Nanotechnology. *Op Acc J Bio Sci & Res* 6(3)-2023 DOI:10.46718/JBGSR.2020.06.000147.
4. Could an artificial intelligence be considered a person under the law? -<https://phys.org/news/2018-10-artificial-intelligence-person-law.html>
5. AI and human creativity go hand in hand - <https://phys.org/news/2018-10-ai-human-creativity.html>
6. Using AI to create new fragrances - <https://phys.org/news/2018-10-ai-fragrances.html>
7. Raj, S.; Jose, S.; Sumod, U.S.; Sabitha, M. Nanotechnology in cosmetics: Opportunities and challenges. *J. Pharm. Bioallied Sci.* **2012**, 4, 186–193. [CrossRef] [PubMed]
8. Kaul, S.; Gulati, N.; Verma, D.; Mukherjee, S.; Nagaich, U. Role of Nanotechnology in Cosmeceuticals: A Review of Recent Advances. *J. Pharm.* **2018**, 2018, 3420204. [CrossRef]
9. Ajazzuddin, M.; Jeswani, G.; Jha, A. Nanocosmetics: Past, Present and Future Trends. *Recent Patents Nanomed.* **2015**, 5, 3–11. [CrossRef]
10. Size, L.B.M.; Report, S.; Size, L.B.M.; Application, B.; Region, B.; Forecasts, S. Get Free, Instant, and Unlimited Access to a PDF Sample Report & Personalized Online Dashboard. 2021; pp. 2021–2028. Available online: <https://main.mohfw.gov.in/sites/default/files/Annual%20Report%202020-21%20English.pdf> (accessed on 15 January 2022).
11. Schneider, G.; Gohla, S.; Schreiber, J.; Kaden, W.; Schönrock, U.; Schmidt-lewerkühne, H.; Kuschel, A.; Petsitis, X.; Pape, W.; Ippen, H.; et al. Connect with Wiley. The Wiley Network. 2021, pp. 2–3. Available online: [https://onlinelibrary.wiley.com/doi/abs/10.1002/14356007.a24\\_219](https://onlinelibrary.wiley.com/doi/abs/10.1002/14356007.a24_219) (accessed on 15 January 2022).
12. Cosmetics—Overview. Available online: <https://www.fda.gov/industry/regulated-products/cosmetics-overview> (accessed on 15 January 2022).
13. Tripathy, S.; Dureja, H. Cosmetics: Regulatory Scenario in USA, EU and India. *J. Pharm. Technol. Res. Manag.* **2015**, 3, 127–139. [CrossRef]
14. Kumar, N.; Kanchan, T.; Unnikrishnan, B.; Thapar, R.; Mithra, P.; Kulkarni, V.; Holla, R.; Bhagwan, D.; Radhakrishnan, Y. Characterization of *Rubia cordifolia* L. root extract and its evaluation of cardioprotective effect in Wistar rat model. *Indian J. Pharmacol.* **2018**, 49, 344–347. [CrossRef]
15. Haryanti, R. Krim Pemutih Wajah dan Keamanannya. *Majalah Farmasetika* **2017**, 2, 5–9. [CrossRef]
16. Search Worldwide, Life-Sciences Literature. 2021, pp. 1–2. Available online: <https://www.cosmeticsandtoiletries.com/regulations>

## Using UML in Software Requirement Analysis Case Study of Academic Organization as an Example.

**Mrs. Anjali R . Shanke(Jadhav)**

M.C.A, M.B.A, M.Phil(Comp Sci), Dept of Management  
VidyaBharati Mahavidyalaya, Amravati  
Email: anjali123shanke@gmail.com

### **Abstract:**

The work presented in this paper is focused on developing a technique of requirement analysis of software engineering part of a software project management. A case study of academic organisation is considered as an example.

A systematic environment for learning is a need of academic organization. Time is ever changing and their rate of change has accelerated to a very high level recently. Social, economic, political as well as cultural changes in the environment are undergoing developments at a higher rate. Academia cannot be isolated from these changes and hence the same is the context for academic organizations.

Requirement analysis of academic organization require thorough investigation regarding the current requirement of academic standards, the problems based in reaching those standards, deriving procedures of quality achievement and major and non-major conformances of the same in academic organizations.

To improve the performance of academic organization there is a need to study the requirement analysis of academic organization. Although requirement analysis is acknowledged as a critical success factor of information system development for organization, mistakes are frequent at the requirement stage. Two of these mistakes are lack of understanding by requirement engineers and miscommunication between user and system analysts. As a result of these problems, information system may not fulfill organizational needs. To prevent these problems, UML models are useful for understanding the problems and communication with people involved in a project

Through these applications some errors will be detected in the existing requirement of an academic organization. Further in this work through the comparison of our proposed method with the conventional inspection, we will be concluding that our method can complement the limitations of the inspection.

**Keywords: Requirement Analysis, UML, Academic, Software**

### **I. Introduction**

Every organization has to utilize resources so as to achieve the organizational goal, which can be accomplished by thorough investigation, .One should know requirement of academic organization.

Requirement research and reflection on practice in last ten years have transformed the field. Both the overall process and the detailed techniques that go into it are far better understood and are known to improve the performance.

The aim of the requirements workflow is

1. To improve performance of academic organization.

2. To determine the needs of the user.
3. To understand the application domain i.e. the environment in which the target software product is to operate
4. The task of the developer at this stage is to determine exactly what the organizations need are and to find out what constraints exist such as deadline, reliability and cost.

The preliminary investigation of the user needs sometimes called concept exploration. The functionality is successively refined and analyzed for technical feasibility and financial justification.

Many UML diagrams of the unified process assist the client in gaining the necessary detailed understanding of what needs to be developed.

On Requirement analysis is the process of understanding the organization needs and expectations from a proposed system or application and is a well defined stage in the software development life cycle model.

The software requirements analysis process covers the complex task of eliciting and documenting the requirements, modeling and analyzing requirements and documenting them as a basis for system design. The requirement analysis function may also fall under the scope of project manager, program manager or business analyst, depending on the organizational hierarchy.

Software requirements analysis and documentation process are critical to software project success. Requirements engineering so an emerging field which deals with the systematic handling of requirements

#### ***A. Requirement analysis in software Engineering***

The Software process is the way we produce software. Different organizations have different software processes. For eg Consider the issue of documentation . Some organizations consider the software they produce to be self documenting , that is , the product can be understood simply by reading the source code. Other organizations , However are documentation intensive . They punctiliously draw up specifications and check them methodically , then they perform design activities, check and recheck their design before coding commences and give extensive descriptions. Once the product has been delivered and installed on the clients computer any suggested change must be proposed in writing , with detailed reasons for making the change.

The preliminary investigation of the client's needs sometimes is called concept exploration. In subsequent meetings between members of the development team and the client team, the functionality of the proposed product is successively refined and analyzed for technical and financial justification. When the product finally is delivered to the user, perhaps a year or two after the specifications have been signed off on by the client, the client may say to the developers, "I know that this is what I asked for, but it isn't really what I wanted , " what the developers thought the client wanted was not what the client actually needed.

To avoid such misunderstanding between the client and developers and to understand the requirements the unified process was been developed. The many UML diagrams of the unified process assist the client in gaining the necessary detailed understanding of what needs to be developed.

## II. Problem definition

The present study is entitled “Study of Requirement analysis using General Purpose Modeling language for academic organization: UML” is taken as a topic of research for following reasons.

To go through the process of software development for an academic organization, the first step is requirement analysis. It has been found through last research that due to improper requirement analysis, many projects fail or need of organization is not fulfilled. Therefore it's necessary to do detailed requirement analysis of an academic organization right from getting proper certification to the student getting admitted to a particular course.

### *A.Problems*

#### *1) Customers don't really know what they want:*

Customers only have a vague idea of what they need. It is necessary to perform requirement analysis to turn this amorphous vision into formally –documented software requirements specification that can, in turn, be used as the basis for both a project plan and an engineering architecture.

#### **Solution:**

- 1 Understand the objectives, deliverables and scope of the project.
- 2 Evaluate risks involved in the project.
- 3 Write a concrete vision statement for the project .
- 4 Both the software engineer and the customer or the user should have the clear understanding of the deliverables.

#### *2).Requirements change during the course of the project*

Requirements of academic organization can change as the project progresses and the user can make necessary corrections .It may also occur because changes in the external environment require reshaping of the original problem. Good software engineer are aware of these possibilities and have backup plans to deal with these changes.

#### **Solution: -**

1. Should have clearly defined process for receiving, analyzing and incorporating change request.
2. Set milestone for each development phase beyond which certain changes are not permissible.
3. Ensure that change request are clearly communicated and the master plan is updated accordingly.

#### *3) Customers have unreasonable timelines.*

A common mistake is to agree to such timelines before actually performing a detailed analysis and understanding both of the scope of the project and the resources necessary to execute it. It 's quite likely that the project will either get delayed or suffer from quality defects .

#### **Solution:**

- 1 Convert the software requirements specification into a project plan, detailing tasks and resources needed at each stage and modeling best case, middle –case and worst-case scenarios.
- 2 Ensure that the project plan takes account of available resource constraints and keeps sufficient time for testing and quality inspection.

#### *4) Communication gap exist between customers/users, engineers and project managers.*

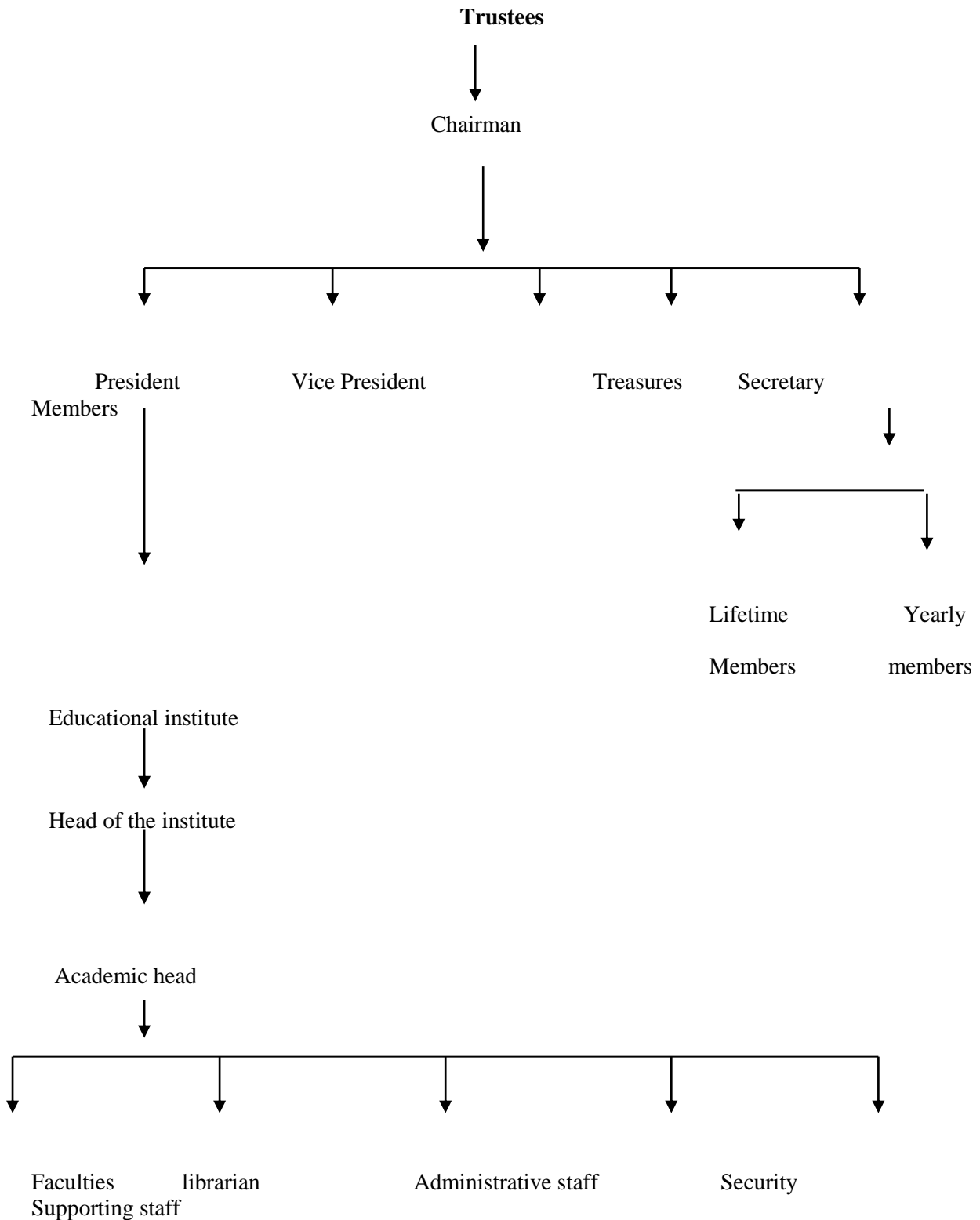
Often customers and engineers fail to communicate clearly with each other because they come from different worlds and do not understand technical terms in the same way. This can lead to confusion and severe miscommunication, and an important task of a project manager; especially during the requirement analysis phase is to ensure that both parties have a precise understanding of the deliverable and tasks needed to achieve it.

#### *5) The development team doesn't understand the politics of the customer's organization*

Understanding user requirements is an integral part of information system design and is critical to the success of interactive systems .It is now widely understood that successful systems and products begin with an understanding of the needs and requirements of the users.

To prevent the problem related with the requirement, UML models are useful for understanding the problem and communication with people involved in a project. Textual requirement specification is difficult to develop, understand, review and maintain- graphical modeling is widely recognized as a more effective analysis tool. UML works as defecto standard in software modeling.

### *Organization chart*



### III. Necessity of UML Model

To improve the performance of academic organization there is a need to study the requirement analysis of academic organization. Although requirements analysis is acknowledged as a critical success factor of information system development for organizations, mistakes are frequent at the requirement stage. Two of these mistakes are lack of understanding by requirement engineers and miscommunication between user and system analysts. As a result of these problems, information system may not fulfil organizational needs. To prevent these problems and communication with people involved in a project.

Communicating the vision is of utmost importance. Before the advent of the UML, system development was often a hit-or-miss proposition. System analysts would try to assess the needs of their clients, generate a requirements analysis in some notation that the analyst understood, give that analysis to a programmer or team of programmers, and hope that the final product was the client wanted.

The power of the unified modeling language is not limited to object oriented software development. More and more, the UML is being applied to other areas of software development, such as data modeling, enhancing practitioners ability to communicate their needs and assessments to the rest of the team.

#### **a Advantages of using UML**

- To provide student with designing software as a team.
- To model the communication problems that are typical in software projects
- To demonstrate how UML helps overcome communication problems.

UML is not restricted to modeling software. UML is also used for business process modeling, System engineering modeling and representing organizations structures. The systems modeling language (SYSML) is a domain –specific modeling language for system engineering that is defined as UML 2.0 profile. UML has been a catalyst for the evolution of model-driven technologies which include model-driven development (MDD), Model driven Engineering (MDE), and Model driven Architecture (MDA). UML has allowed software developers to concentrate more on design and architecture.

UML diagrams for requirement analysis for Academic organization

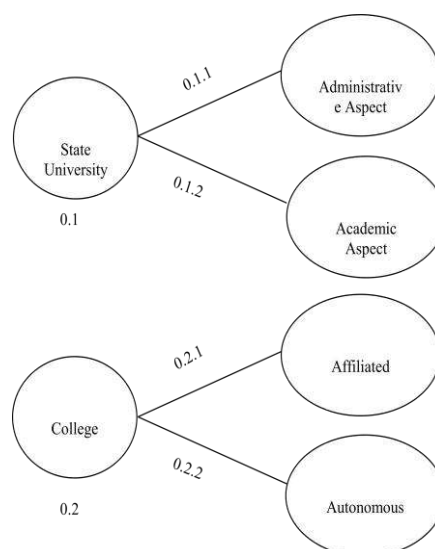


Fig 1 Level 0 UML Diagram

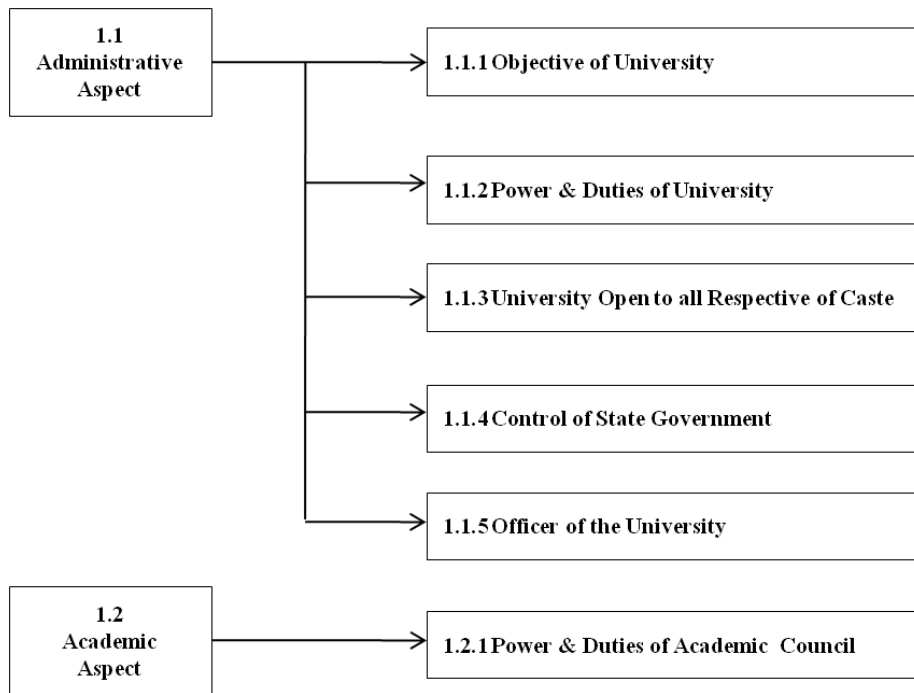
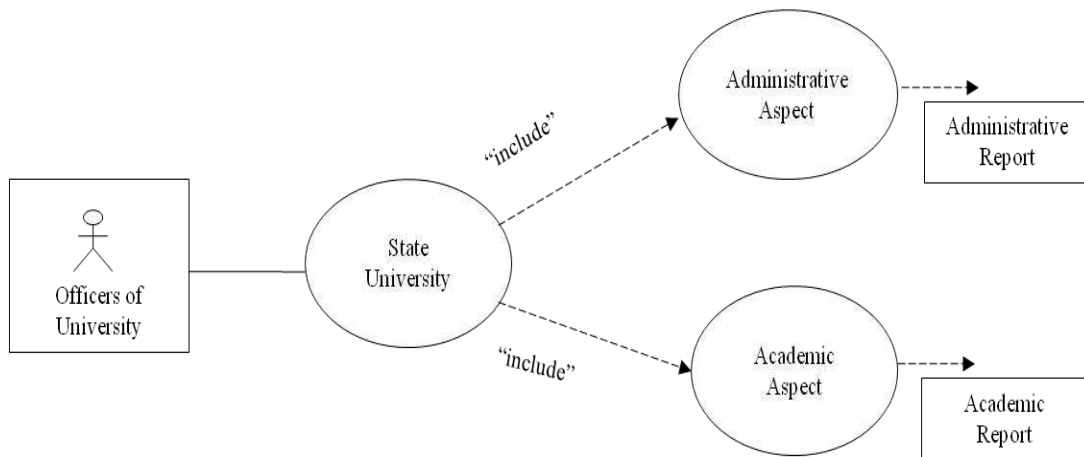


Fig 2. Level 1 UML Diagram Of State University



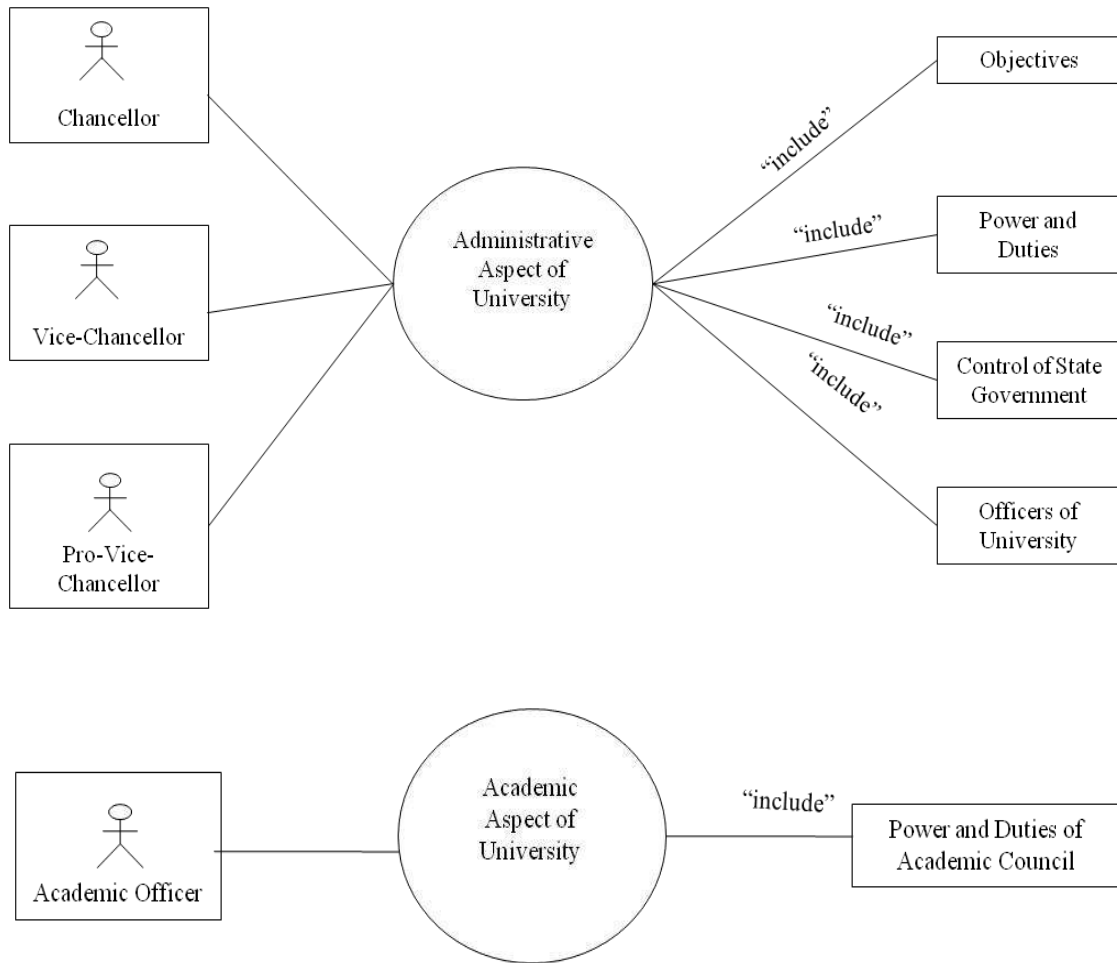


Fig 3 Use Case Diagram

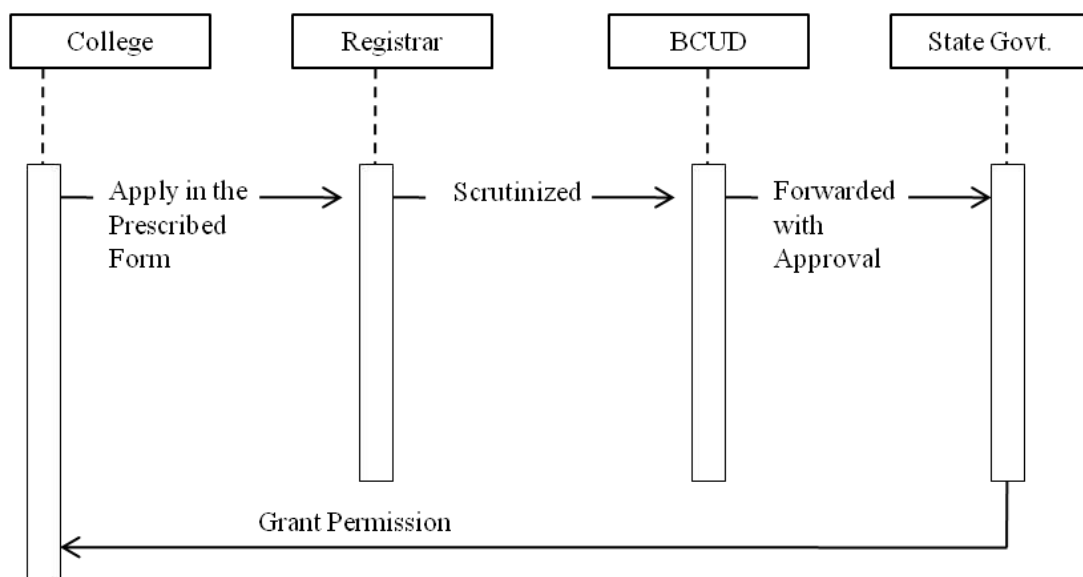




Fig 4 Procedure for Permission

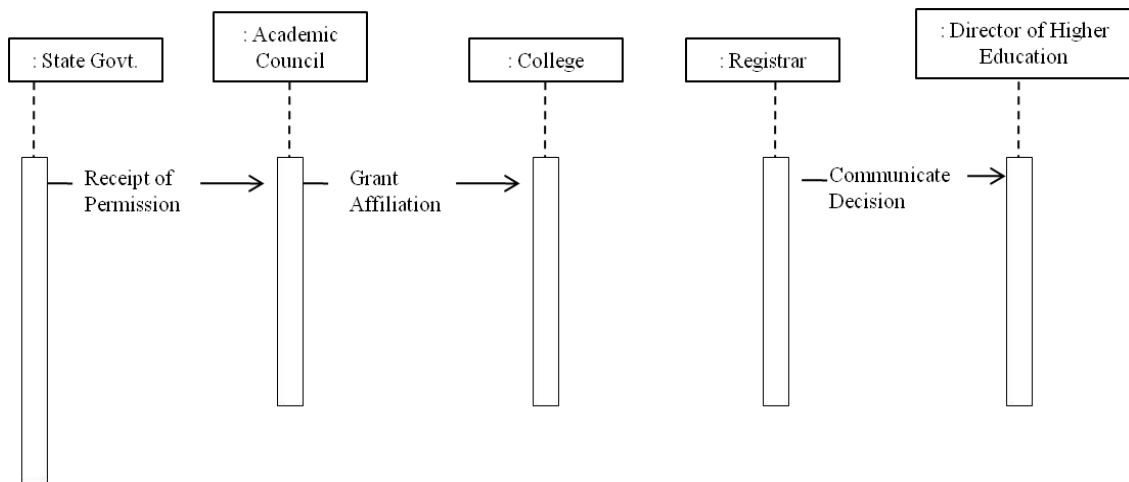


Fig 5. Procedure for Affiliation

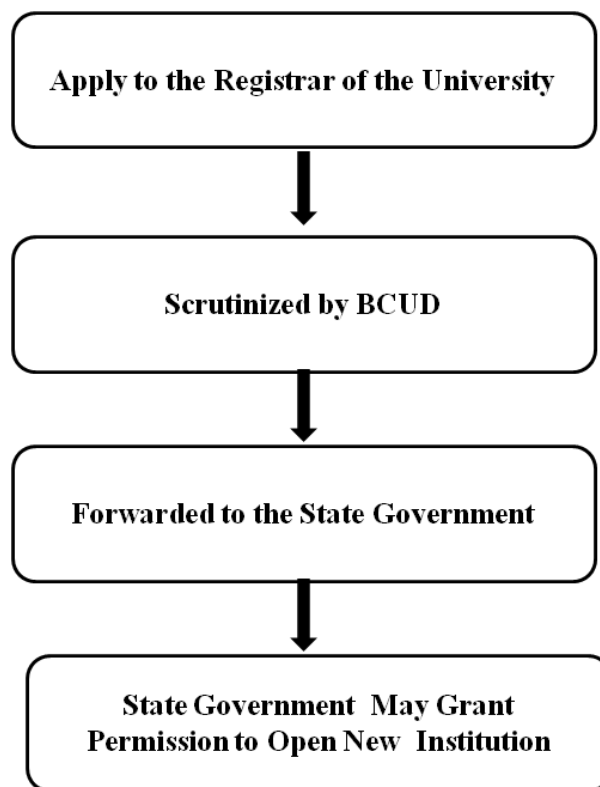


Fig 6 Activity Diagram

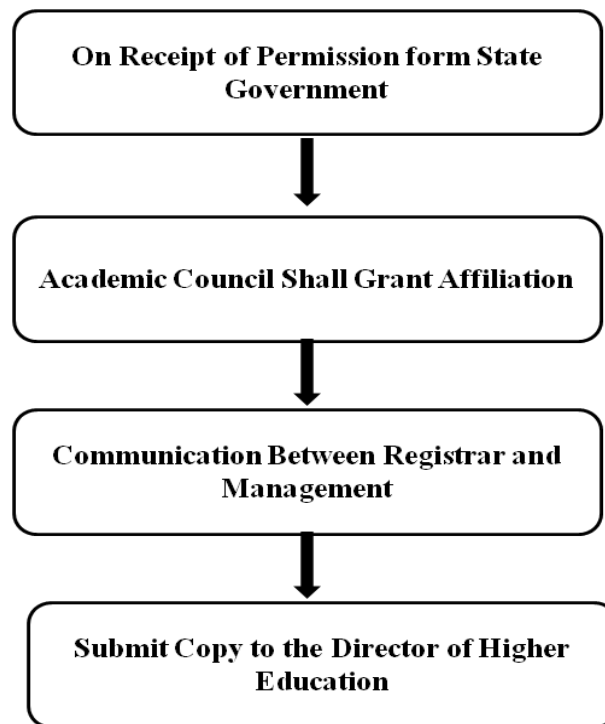


Fig 7 Activity Diagram

#### IV. Conclusion

The work presented here carries out the requirement analysis of academic organization considering a modeling process of requirement engineering, which is a part of software engineering project management.

The requirement analysis of the organization is carried out in this dissertation considering current requirement of academic standards. The problems based in teaching the standards, deriving procedures of quality achievement and major issues in academic organization. The requirement analysis carried out is presented in form of standard model using UML.

The UML representation then used to develop software model. The process indicates the way of successfully carrying out of requirement analysis of any organization through UML modeling and converting it to the successful software module.

#### V. Limitation

The process of work carried out in this dissertation indicates that for any organization requirement analysis must be carried out through understanding of the needs and expectations of the proposed system. It must cover complex task and documenting the requirement of the organization.

#### Scope

1. Requirement analysis and the requirement engineering process are the important steps from software engineering which leads to successful software project management.
2. Requirement engineering is an engineering field which deals with the systematic handling of requirements.
3. It can be carried out for any type of organization for successful implementation of software project management.

## VI. References

- [1] G.Booch, J.Rambaugh, I.Jacobson, *the unified modeling language user guide*,Addison-Wesley,1998.ISBN 0-201-57168-4.
- [2] J.Rambaugh, I.Jacobson, G.Booch,*the unified modeling language reference manual*,Addison-wesley,1998. ISBN 0-201-30998.
- [3] I.Jacobson, G.Booch, j. Rambaugh, *the unified software development process*, Addison-Weseley, 1999 ISBN-201-65783.
- [4] The object management group home of the UML specification [www.omg.org](http://www.omg.org)
- [5] S.Shlaer & S Mellor, *object-oriented system analysis modeling the world in data*, yourdan press, 1989.ISBN 0-13-629023
- [6] R.Young, *effective requirements practices*, Addison-wesley, 2001,ISBN 0-201-70912-0
- [7] Graham, L.Graham, *Requirements engineering & Rapid development. An object –oriented Approach*, Addison – Wesley, 1998. ISBN 0-201-36047-0
- [8] S.Robertson, J. Robertson, *Mastering the requirements process*, Adisson-wesley,2000. ISBN 0-201-657678
- [9] *object –oriented system analysis and design using UML* by Bennett
- [10] *Object –oriented software engineering :using UML ,patterns and java* by Bernd Bruegge ,Allen H. Dutoit.
- [11] *System requirement analysis* ,by Jefferey O.Grady
- [12] *Requirement analysis & system design-* Leszek Maciazek

## Artificial Intelligence and it's Applications

**Prof. Pratik S. Yawale**

Assistant Professor, CSE Dept., P.R. Pote College of Engineering & Management, Amravati.

**Prof. Devendra G. Ingale**

Assistant Professor, CSE Dept., Dr. Rajendra Gode Institute of Technology & Research, Amravati.

### ***Abstract:***

Artificial Intelligence (AI) is truly a revolutionary branch of computer science, set to become a core component of all modern software over the coming years and decades. This creates a threat, but also an opportunity. AI will be deployed to augment both defensive and offensive cyber operations. Additionally, new ways of cyber-attack are invented to take advantage of the particular weaknesses of AI technology. AI is broadly characterized as the study of computations that allow for perception, reason and action. Today, the amount of data generated, by both humans and machines, far outpaces humans ability to absorb, interpret, and make complex decisions based on that data. This paper examines features of Artificial Intelligence, applications, growth and achievements.

***KEYWORDS- AI, Machine Learning, Deep learning, Neural networks, Natural Language Processing, Automation and Robotics***

### **INTRODUCTION:**

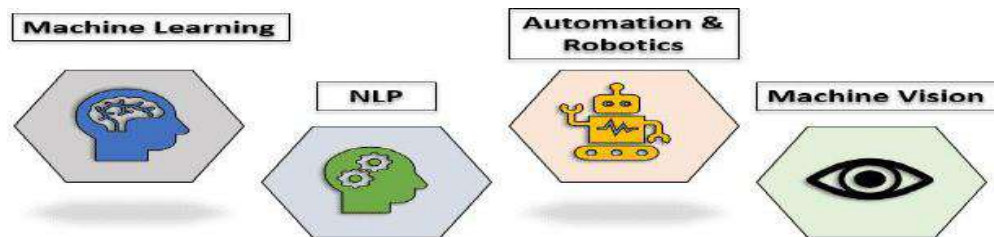
In general terms, AI refers to computational tool that is able to substitute for human intelligence in the performance of certain tasks. This technology is currently advancing at a breakneck pace, much like the exponential growth experienced by database technology in the late twentieth century. The central principle of AI includes reasoning, knowledge, planning, learning, communication, perception and the ability to move and manipulate objects. It is the science of making intelligent machines, especially intelligent computer programs. One of the essential purposes of AI is to automate tasks that previously would have required human intelligence.

### **ARTIFICIAL INTELLIGENCE METHODS:**

**Machine Learning** is one of the applications of AI where machines are not explicitly programmed to perform certain tasks; rather, they learn and improve from experience automatically. Deep Learning is a subset of machine learning based on artificial neural networks for predictive analysis. There are various machine learning algorithms, such as Unsupervised Learning, Supervised Learning, and Reinforcement Learning. In Unsupervised Learning, the algorithm does not use classified information to act on it without any guidance. In Supervised Learning, it deduces a function from the training data, which consists of a set of an input object and the desired output. Reinforcement learning is used by machines to take suitable actions to increase the reward to find the best possibility which should be taken in to account.

**Natural Language Processing (NLP):** It is the interactions between computers and human language where the computers are programmed to process natural languages. In NLP, the audio of a human talk is captured by the machine. Then the audio to text conversation occurs, and then the text is processed where the data is converted into audio. Then the machine uses the audio to respond to humans. Applications of Natural Language Processing can be found in IVR

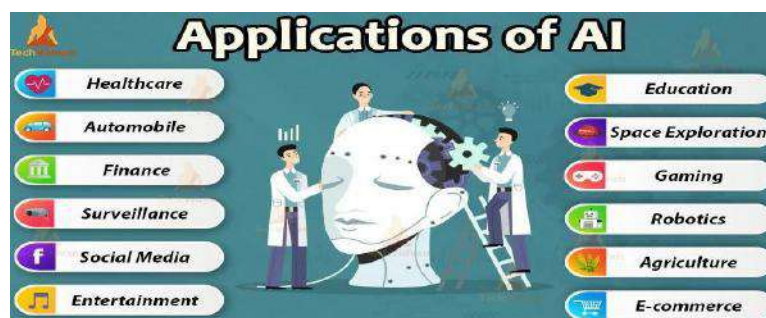
(Interactive Voice Response) applications used in call centres, language translation applications like Google Translate and word processors such as Microsoft Word to check the accuracy of grammar in text.



**Automation and Robotics:** Automation aims to improve productivity and efficiency by having machines perform monotonous and repetitive tasks which results in cost-effective outcome. Organizations use machine learning, neural networks, and graphs in automation. Using CAPTCHA technology, such automation prevents fraud issues during online financial transactions. Programmers create robotic process automation to perform high-volume repetitive task that can adapt changes in different circumstances.

**Machine Vision:** Machines can capture visual information and then analyze it. This process involves using cameras to capture visual information, converting the analog image to digital data, and processing the data through digital signal processing. The usage of machine vision can be found in signature identification, pattern recognition, medical image analysis, etc.

### APPLICATIONS OF ARTIFICIAL INTELLIGENCE:



**1. Healthcare:** One of the foremost deep-lying impacts which AI has created is within the Healthcare space. A device, as common as a Fitbit or an iWatch, collects a lot of data like the sleep patterns of the individual, the calories burnt by him, heart rate and a lot more which can help with early detection, personalization, even disease diagnosis. This device, when powered with AI can easily monitor and notify abnormal trends. This can even schedule a visit to the closest Doctor by itself and therefore, it's also of great help to the doctors who can get help in making decisions and research with AI.

**2. Automobile:** At this stage where automobiles changing from an engine with a chassis around it to a software-controlled intelligent machine, the role of AI cannot be underestimated. The goal of self-driving cars, during which Autopilot by Tesla has been the frontrunner, takes up data from all the Tesla's running on the road and uses it in machine learning algorithms. The assessment of both chips is later matched by the system and followed if the input from both is the same.

**3. Banking and Finance:** One of the early adopter of Artificial Intelligence is the Banking and Finance Industry. From Chatbots offered by banks, for instance , SIA by depository financial institution of India, to intelligent robo-traders by Aidya and Nomura Securities for

---

autonomous, high-frequency trading, the uses are innumerable. Features like AI bots, digital payment advisers and biometric fraud detection mechanisms cause higher quality of services to a wider customer base.

**4. Social Media:** All of us love Social Media. Social Media is not just a platform for networking and expressing oneself. It subconsciously shapes our choices, ideologies, and temperament. All this due to the synthetic Intelligence tools which work silently within the background, showing us posts that we “might” like and advertising products that “might” be useful based on our search and browsing history. For example, recently Instagram revealed how it’s been using AI to customize content for the Explore Tab. This helps with social media advertising because of its unprecedented ability to run paid ads to platform users based on highly granular demographic and behavioral targeting.

**5. Education:** In the education sector also, there are a number of problems which will be solved by the implementation of AI. A few of them being automated marking software, content retention techniques and suggesting improvements that are required. This can help the teachers monitor not just the academic but also the psychological, mental and physical well being of the students but also their all-round development. This would also help in extending the reach of education to areas where quality educators can’t be present physically.

**6. Space Exploration:** AI systems are being developed to scale back the danger of human life that venture into the vast realms of the undiscovered and unraveled universe which is a very risky task that the astronauts need to take up. As a result, unmanned space exploration missions just like the Mars Rover are possible due to the utilization of AI. It helps us to discover numerous exoplanets, stars, galaxies, and more recently, two new planets in our very own system. NASA is also working with AI applications for space exploration to automate image analysis and to develop autonomous spacecraft that would avoid space debris without human intervention, create communication networks more efficient and distortion-free by using an AI-based device.

**7. Robotics:** With increasing developments within the field of AI, robots are becoming more efficient in performing tasks that earlier were too complex. The idea of complete automation are often realized only with the assistance of AI, where the system can’t just perform the specified task but also monitor, inspect and improve them without any human intervention. AI in robotics helps the robots to learn the processes and perform the tasks with complete autonomy, without any human intervention.

**8. Agriculture:** Artificial Intelligence is changing the way we do one among our most primitive and basic professions which is farming. The uses of AI in agriculture are often attributed to agriculture robots, predictive analysis, and crop and soil monitoring. In addition, drones are also used for spraying insecticides and detecting weed formation in large farms. This is getting to help firms like Blue River Technologies, better manage the farms. AI has also enhanced crop production and improved real-time monitoring, harvesting, processing and marketing.

### **FUTURE OF ARTIFICIAL INTELLIGENCE:**

In the last year or so, a subset of AI—generative AI—has been gaining traction. Generative AI uses deep learning to analyse existing sets of data to create new outputs. Unlike its predecessors, generative AI also has reasoning capabilities. ChatGPT, which can produce human-like responses to text prompts, and DALL- E, which can create images and artworks from text prompts, are popular examples of generative AI. The rise of generative AI has raised curiosity and piqued interests. It feels like the future of AI is a rapidly changing landscape,

that's because the present innovations in the field of artificial intelligence are accelerating at such a blazing-fast pace that it's tough to keep up. Indeed, artificial intelligence is shaping the future of humanity across nearly every industry.

### **CONCLUSION:**

Till now we have discussed in brief about Artificial Intelligence. We have discussed some of its principles, its applications, its achievements etc. AI leads to transformative applications within a series of industrial, intellectual, and social applications, far beyond those caused by previous industrial revolutions. Furthermore, AI has proven to be superior to human decision-making in certain areas. AI is better than humans at finding and enacting the best policies in certain areas concerning science, engineering, and complex societal and macroeconomic issues.

### **REFERENCES:**

- [1] [https://idhjournal.com/article/S2468-0451\(18\)30144-5/fulltext](https://idhjournal.com/article/S2468-0451(18)30144-5/fulltext)
- [2] <https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c>
- [3] <https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=453e91f0233e>
- [4] <https://www.indiatoday.in/education-today/news/story/robots-teachers-bengaluru-school-artificial-intelligence-ai-divd-1594366-2019-09-02>
- [5] <https://www.exametc.com/magazine/details.php?id=521>

## Cyber Security Awareness on Cyber Attacks

**Aparna R.Sapate.**

Assi.prof .Dept.of comp. Sci.  
VidyaBharti Mahavidyalaya, Amravati  
Email:aparnasapate268@gmail.com

### **Abstract:**

Now, a day we know very well the electronic technology is very important. That's why today's world is highly and deeply dependent on electronic technology, which are many types of devices, and mostly important role of that is information technology in field of cyber security. That is plays crucial work to secure cyber security and protecting the data from cyber-attacks because it is a challenging issue. Cyber security is an important role of information technology for securing the data Systems, important files, data, and other important virtual things are at risk if there is no security to protect it. When ever we think about the cyber security the first thing that comes to our mind to protect the data From 'cyber attack' which are increasing extremely day by day. various types of companies and Government organisations are work continuously after taking many measures in order to prevent these cyber crimes. That's why Focuses on Challenges faced by cyber security on the latest technologies. And we should have to focused the latest cyber security trends, techniques and also ethics changing the face of cyber security. but to prevents the cyber attacks Besides various measures cyber security is still actually it is a very big concern to many. Otherwise military, government, financial, medical and corporate organizations accumulate, practise, and stock unprecedented quantities of data on PCs are essential in cyber security and other devices. It's all of the important quota which can be very sensitive information.

**Keywords:** cyber security, cyber attacks, cyber crime, cyber ethics, social media, cloud computing.

### **Introduction:**

Today cyber security is the process of protecting Systems, network and program from digital attacks. it's used to accessing, changing, or destroying sensitive information also cyber security refers to every aspect of protecting an organization and exporting money from users via ransomware; or interrupting normal business processes also employees and assets against cyber threats. cyber Security is crucial also security awareness is the process of educating people to understand, identify that how to protect and stay security and avoid the cyber threats. the cyber Security of today, work on defending computer, network, database and smartphones from the threat. most people just don't have knowledge, tools and support they need to protect themselves and their organization. And the average person's cybersecurity knowledge. [4] The youth in the present world have embraced the internet based communication methods faster than the elders and they are now at a state where they can't even imagine a world without internet and smart phones. These devices have now become a part of their daily life and they spend a considerable amount of time using computers, smart phones and especially social media. Hence, there is a huge threat to youth, Therefore it is mandatory to analyze the awareness level of cyber security among the youth. [3] cybersecurity outbreak can result in entirety from individuality theft, cyber attacks to blackmail attempts, to the damage of vital data similar family photographs. Everybody relies on dangerous structure like influence plants, infirmaries, and monetary service businesses. Securing these and other societies is essential to trust our civilization operative. There are many varieties of cyber attacks that happen in the world today. If we know



the various types of cyberattacks, it becomes easier for us to protect our networks and systems against them. Here, we will closely examine the top ten cyber-attacks that can affect an individual, or a large business, depending on the scale. The fight against cyber crime needs a comprehensive and safer approach. Given that technical measures alone cannot prevent any crime, it is critical that law enforcement agencies are allowed to investigate and prosecute cyber crime effectively.[1]Hence cyber security has become a latest issue. The scope of cyber security is not just limited to securing the information in IT industry but also to various other fields like cyber space etc.

### **Cyber Security:**

Cyber security is the process of protecting Systems, network, data and program from digital attacks. It's used to accessing, changing, or destroying sensitive information also cyber security refers to every aspect of protecting an organization and exporting money from users via ransomware or interrupting normal business processes also employees and assets against cyber threats. Privacy and security of the data will always be top security measures that any organization takes care. We are presently living in a world where all the information is maintained in a digital or a cyber form. Social networking sites provide a space where users feel safe as they interact with friends and family. In the case of home users, cyber-criminals continuously target to social media sites to steal personal data. Not only social networking but also during bank transactions a person must take all the required security measures. As cyber attacks become more common and corporate networks grow more complicated, implementing effective cyber security measure is particularly challenging today and a variety of cyber security solutions are required to mitigate corporate cyber risk and also that process.

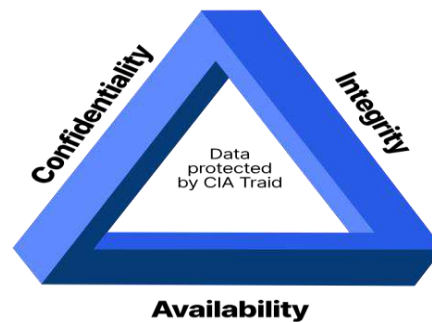
**People:** Users must understand and comply with basic data security principles like choosing strong passwords, being wary of attachments in email, and backing up data. Learn more about basic cybersecurity principles with these **1.** Keep personal information private, **2.** Use caution to avoid bad actors, **3.** Update software regularly, **4.** Create strong passwords by using paraphrases, **5.** Use two-step verification whenever possible, **6.** Be cautious of free Wi-Fi, **7.** Don't leave a cyber footprint on shared or public devices, **8.** Manage your privacy settings, **9.** Regularly audit applications you have installed as privacy settings can change with upgrades, **10.** Secure tomorrow, 10 together

**Processes:** Organizations must have a framework for how they deal with both attempted and successful cyber attacks. One well-respected framework can guide you. It explains how you can identify attacks, protect systems, detect and respond to threats, and recover from successful attacks. Learn about the the NIST cybersecurity framework. NIST is The Cybersecurity Framework (CSF) is a set of cybersecurity best practices and recommendations from the National Institute of Standards and Technology (NIST). The CSF makes it easier to understand cyber risks and improve your defenses.

**Principles of Cyber Security:** The primary objective of cyber security is to ensure protect data. The security community commonly refers to a triangle of three related principles that ensure data is secure, known as the CIA triad:

- **Confidentiality** — ensuring sensitive data is only accessible to those people who actually need it, and are permitted to access according to organizational policies, while blocking access to others.
- **Integrity** — making sure data and systems are not modified due to actions by threat actors, or accidental modification. Measures should be taken to prevent corruption or loss of sensitive data, and to speedily recover from such an event if it occurs.
- **Availability** — ensuring that data remains available and useful for its end-users, and that this access is not hindered by system malfunction, cyber attacks, or even security measures themselves. The CIA Triad defines three key principles of data security .

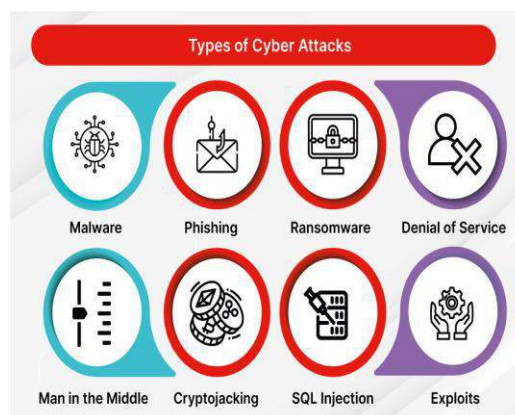
To achieve the CIA objectives organizations must protect two aspects of their IT environment: application security and data security.



**Fig.1 Principles of Cyber Security:**

### Cyber Attacks:

The most important cyber-attacks methods are Denial of service, logical bomb, Abuse tools, Sniffer, Trojan horse, Virus, Worm, Send spam, and Botnet. Fig. 1. illustrates the important cyber-attacks types. In the Denial of service method, the authorized users access to the system and vice versa is lost. In fact, the attacker from one point starts immersing the target computers in various messages and blocking the legal flow of data. This prevents any system from using the Internet or communicating with other System.[6] Cyber has increased the yield of the community and effectively distributed information over time. No problem what application or industry cyber is used in, increasing production has always been considered. Fast data transfer to cyberspace mostly declines the total system security. Cyber crime is a term for any illegal activity that uses a computer as its primary means of commission and theft. such as network intrusions and the dissemination of computer viruses, as well as computer-based variations of existing crimes, such as identity theft, stalking, bullying and terrorism which have become as major problem to people and nations. As day by day technology is playing in major role in a person's life the cyber crimes also will increase along with the technological advances.[3] Cyber Security Privacy and security of the data will always attention



**Fig .2 Types of cyber Attacks:**

## Types of cyber Attacks:

**1. Malware:** Malware is a form of application that performs nefarious activities. Some types of malware are designed to create access to networks, some to spy on credentials while others are simply used to cause disruption.

Malwares can be used for extortion as well. An example of it can be found in Ransomware attacks of 2017 where a program was designed to encrypt the victim's files and then ask them to pay a ransom in order to get the decryption key.

**Latest Update about Malware Attacks in India** :-Mobile Security Report 2021 asserted that mobile malware attacks in India are on rise (845 percent increase) since October 2020.

**2. Phishing:** In Phishing, an attacker tricks an unsuspecting target into handing over valuable information, such as passwords, credit card details, etc. An example of this is a message regarding One-Time Passwords (OTP). A hacker using a phishing method will send a clickable link where a user can submit their OTPs. Once the link is clicked a hacker will have access to the users personal information. Phishing is the common form of cyber attack due to its effectiveness and simplistic execution pattern.

**Latest Update about Phishing in India**:-Indian Computer Emergency Response Team (CERT-In) released a public advisory to alert citizens against all attempts of phishing through fake domains, emails and text messages that promise registration for a job against the pandemic.

**3. Man-in-the-middle attack (MITM):** A man-in-the-middle attack (MITM) consists of a message interception between two parties in an attempt to spy on the targets. Due to the advent of end-to-end encryption, MITM attacks have taken a dip in frequency of attacks. Such encryptions prevent third parties in intercepting or tampering data transmitted in the network. Whether the network is secure or not is hardly a factor.

**4. Distributed Denial-of-Service (DDoS) attack:** In a DDoS attack, an attacker floods a target server with traffic that will disrupt it. Since most servers cannot handle it, it may lead to services slowing down on the website and if it eventually crashes. Unlike standard denial-of-service attacks, DDoS uses multiple compromised devices to bombard the target server, which sophisticated firewalls cannot respond to or are unable to.

**Update about Distributed Denial-of-Service attack in India**:-In August 2020 the number of Distributed Denial of Service (DDoS) incidents in India hit a record high in terms of total DDOS packets, which were well in excess of 10 billion as per a study by global cyber security firm Radware

**5. SQL Injection:** This type of cyber attack targets specific SQL databases. These databases use SQL statements for data query. In case permissions are not set properly, a hacker can manipulate SQL queries into changing the data if not deleting them altogether.

**6. Zero-day exploit:** When cyber-criminals learn of a vulnerability in a frequently used software application they target users and organizations using the software to exploit it until a fix is available. This is called a Zero-day exploit.

**7. DNS Tunnelling:** A DNS Tunnelling provides attackers with a stable and consistent line of communication to the given target. The malware used will gather information as long as the DNS tunnelling is active. Chances are that firewalls won't be able detect such an attack.

**Update about DNS Tunneling in India**:-India saw the highest number of domain name system or DNS attacks in 2020 with 12.13 attacks per organisation, even though the cost of attacks in the country decreased by 6.08% to ₹5.97 crores, said International Data Corporation or IDC's DNS Threat Report.

**8. Business Email Compromise (BEC):** In a BEC attack, hackers target employees who have specific authority to finalize business transactions. They trick them into transferring money into an account belonging to the hacker.

BEC attacks are the most common, if not one of the most damaging attacks for a business firm.

**9. Cryptojacking:** Cryptojacking is used to target a computer in order to mine cryptocurrencies such as bitcoin. The hackers will be able to get all the cryptocurrency they can instead of the original owners. Cryptojacking is not so widely known but its severity cannot be underestimated.

**10. Drive-by Attack:** A website is loaded with a malware, and when a visitor happens to come across such a website their device is infected with the malware. The malware will steal valuable data or crash the system.[8]

### **Conclusion:**

The target of the report was to present the importance of Cyber Security in India and also analyze the present framework and policies of Government. The study concludes that what is cyber Security, and what is cyber attacks and that types, it has been an exponential increase in the number of Cyber Attacks throughout the globe causing large amount of loss to their system. The response to the Cyber Attacks but such policies failed due to lack of resources and serious government interest. As a result the users in India are still not safe when connected to the Internet. Each new each day, are challenging organizations with not only how they secure their infrastructure, but how they require new platforms and intelligence to do so. There is no perfect solution for cyber-crimes but we should try our level best to minimize them in order to have a safe and secure future in cyber space. They should be aware on how to protect our privacy management and encrypt their personal data, which is still in a very low level. In sum, the government should be alarmed of this situation and should implement relevant laws and regulations, even after they themselves were hacked during the pandemic in last few months. Measures should be taken to enhance cyber security in the country and further delay will lead to unbearable circumstances.[9]

### **References:**

1. A Study Of Cyber Security Challenges And Its Emerging Trends On Latest Technologies February 2014  
[https://www.researchgate.net/publication/260126665\\_A\\_Study\\_Of\\_Cyber\\_Security\\_Challenges\\_And\\_Its\\_Emerging\\_Trends\\_On\\_Latest\\_Technologies](https://www.researchgate.net/publication/260126665_A_Study_Of_Cyber_Security_Challenges_And_Its_Emerging_Trends_On_Latest_Technologies).
2. S. Tayal, N. Gupta, P. Gupta, D. Goyal, and M. Goyal, "A Review paper on Network Security and Cryptography." *Advances in Computational Sciences and Technology*, vol. 10 (5), pp. 763-770, 2017.
3. R. Khan, and M. Hasan, "Network threats, attacks and security measures: A review." *International Journal of Advanced Research in Computer Science*, vol. 8 (8), pp. 116-120, 2017.
4. S. Rathore, P. K. Sharma, V. Loia, Y.-S. Jeong, and J. H. Park, "Social network security: Issues, challenges, threats, and solutions." *Information Sciences*, vol. 421 pp. 43-69, 2017.
4. [4] S. Gao, Z. Li, B. Xiao, and G. Wei, "Security threats in the data plane of software-defined networks." *IEEE Network*, vol. 32 (4), pp. 108-113, 2018.
5. [5] P. Sinha, A. Kumar Rai, and B. Bhushan "Information Security threats and attacks with conceivable counteraction," In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT). IEEE, pp. 1208-1213, 2019.
6. A. Tayal, N. Mishra, and S. Sharma, "Active monitoring & postmortem forensic analysis of network threats: A survey." *International Journal of Electronics and Information Engineering*, vol. 6 (1), pp. 49-59, 2017.

7. M. Wazid, A. K. Das, V. Odelu, N. Kumar, M. Conti, and M. Jo, "Design of secure user authenticated key management protocol for generic IoT networks." *IEEE Internet of Things Journal*, vol. 5 (1), pp. 269-282, 2017.
8. H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements." *IEEE access*, vol. 5pp. 1872-1899, 2017.
9. D. Barrera, I. Molloy, and H. Huang "Standardizing IoT network security policy enforcement," In: *Workshop on Decentralized IoT Security and Standards (DISS)*. p 6, 2018.
10. S. Zheng, Z. Li, and B. Li "Implementation and application of ACL in campus network," In: *AIP Conference Proceedings*. vol 1. AIP Publishing LLC, p 090014, 2017.
11. T. Hayajneh, S. Ullah, B. J. Mohd, and K. S. Balagani, "An enhanced WLAN security system with FPGA implementation for multimedia applications." *IEEE Systems Journal*, vol. 11 (4), pp. 2536-2545, 2015.
12. P. Sinha, V. Jha, A. K. Rai, and B. Bhushan "Security vulnerabilities attacks and countermeasures in wireless sensor networks at various layers of OSI reference model: A survey," In: *2017 International Conference on Signal Processing and Communication (ICSPC)*. IEEE, pp. 288-293, 2017.
13. V. Pruthi, K. Mittal, N. Sharma, and I. Kaushik "Network Layers Threats & its Countermeasures in WSNs," In: *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, pp. 156-163, 2019.
14. J. Singh, Y. Bello, A. Refaey, and A. Mohamed, "Five-Layers SDP- Based Hierarchical Security Paradigm for Multiaccess Edge Computing." *arXiv preprint arXiv :200701246*, vol. pp., 2020.
15. N. Wagner, C. Ş. Şahin, J. Pena, J. Riordan, and S. Neumayer "Capturing the security effects of network segmentation via a continuous-time markov chain model," In: *Proceedings of the 50th Annual Simulation Symposium*. Society for Computer Simulation International, p 17, 2017.
16. M. Oqaily, Y. Jarraya, M. Mohammady, S. Majumdar, M. Pourzandi, L. Wang, and M. Debbabi, "SegGuard: Segmentation-based Anonymization of Network Data in Clouds for Privacy-Preserving Security Auditing." *IEEE Transactions on Dependable and Secure Computing*, vol. pp., 2019

## IOT Based Vehicle Speed Control by Using Mobile

**Prof. Devendra G. Ingale<sup>\*1</sup>, Prof. Pratik S. Yawale<sup>\*2</sup>**

<sup>\*1</sup>Assistant Professor CSE Department Dr.Rajendra Gode Institute of Technology and Research, Amravati.

<sup>\*2</sup> Assistant Professor CSE Department P.R.Pote College of Engineering & Management, Amravati.

### ABSTRACT

Now a days people are driving very fast, accidents are occurring frequently, we lost our valuable life by making small mistake while driving. Accidents are occurring frequently in highly traffic areas. Drivers drive vigorously without caring the traffic. Intimation of driver about speed and accident prone zone is necessary. It can be done by using IOT technology with the help of sensors. This Research is all about controlling speed of vehicle remotely using mobile. This Research can be used to avoid the rash driving of the drivers in restricted areas such schools, parks, hospitals and in speed limited areas etc. The mobile operated vehicle is a concept where a human being can control a vehicles speed by an android app or wireless operation, without physically being sated it. Android app which will be connected to this system by Cloud Server using the Node MCU ESP-32S" is switched on one can operate the vehicle by wireless commands given from app.

### I) INTRODUCTION

The report will cover the development of the vehicle speed detecting and controlling by using mobile application. As the days of manned driving are getting extremely numbered, so are those of traffic jams, bad, dangerous and rough drivers and more importantly, accidents. Automation of the driving control of vehicles is one of the most vital need of the hour. This technology can very well implement what was absent before, controlled lane driving.

Considering the hazards of driving and their more pronounced effect on vehicles our "**IOT Based Vehicle Speed Control by Using Mobile**" is exactly what is required.

Safety is a necessary part of man's life. Due to the accident cases reported daily on the major roads in all parts of the developed and developing countries, more attention is needed for research in the designing an efficient car driving aiding system. It is expected that if such a device is designed and incorporated into our vehicles as a road safety device, it will reduce the incidence of accidents on our roads and various premises, with subsequent reduction in loss of life and property.

### II) **PROBLEM DEFINITION**

- There are unsafe factors

Traffic accidents occur for various reasons. While problems with roads or safety facilities lead to some accidents, the majority of traffic accidents are caused by drivers' failure to abide by regulations, consider pedestrians, and acknowledge dangerous behaviors.

- **Unsafe road environments**

Unsafe road environments refer to external factors uncontrollable by drivers, such as visibility impairment by darkness, slippery surface, insufficient safety facilities, inadequately repaired vehicles, pedestrians or other vehicles that suddenly get in the way.

➤ **Insufficient driver knowledge**

Traffic accidents are often caused by ignorance. Most driver knowledge is acquired through experience. This is why so many new circumstances lead to accidents. If you know what happens when you speed or suddenly stop under special circumstances including rain, snow or a winding road, you would be careful not to speed or brake suddenly.

➤ **Failure to recognize danger**

While there are some drivers who slow down upon recognizing potential dangers of certain situations, others do not see any possible peril. These differences in danger recognition stem from experiences and, in particular, different standards. Drivers with stronger desire to arrive at their destination as soon as possible are more likely to take risks.

➤ **Improper thinking**

There are many types of improper thinking that lead to reckless driving. Such thinking includes believing that it is ok to violate traffic regulations as long as you do not cause accident; rushing to get to your destination even when you are not late; and regarding pedestrians on the road as obstacles.

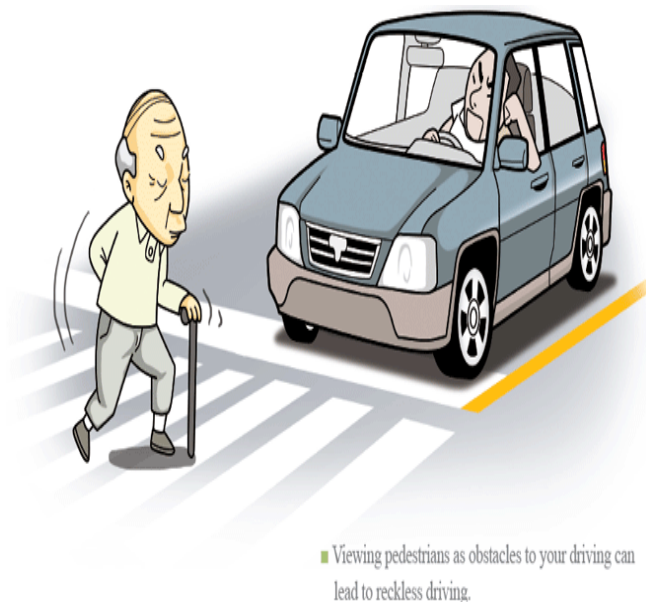


Fig.1.1 Wrong Driving Habits

➤ **Wrong driving habits**

Many drivers tend to wrongly believe that their undesirable driving habits do not pose any danger just because they have yet to cause traffic accidents. This belief can reinforce such habits and thus leads to fatal accidents.

### III) EXISTING SYSTEM

Presently different types of vehicle speed limiters are in use for regulating traffic, especially across roads near populated areas such as hospitals, schools and colleges. There are

many existing systems are coming over this problem. The existing systems are based on speed cameras, RF transmitter and receiver, alarm system or speed limitation devices.

Speed cameras in critical zoned areas only able to capture the high-speed vehicle. It will not reduce the speed of the vehicle and thus it cannot reduce the accident [3]. The implementation of RF transmitter and receiver has some disadvantages such as when a greater number of vehicles crosses the RF devices the interference of signals creates problems. Initial Cost is also high as exclusive RF devices required for every critical zone and areas. RF devices are removable, and it can be removed by any other persons. The Speed of the vehicle does not reduce in the critical zones like colleges and school. To overcome this problem a project aimed to implement an automatic speed limiting system in critical zone is presented. An Android application has been developed with the GPS based system in order to identify the location automatically. When the vehicle reaches the critical zone, the GPS device transmits the signal to the receiver hardware through Bluetooth.[5] The android device shows critical zone areas and indicates the notification to slow down the vehicle speed. When the vehicle reaches the particular zone, the GPS device transmits the signal to the receiver hardware by the Bluetooth. The mechanism associated with the system automatically reduced the speed of the vehicle. This project presents a novel method by which vehicle speed is controlled automatically. The speed measurement and control is accomplished via microcontroller with the signals being received wirelessly from GPS. Apart from its implementation in human operated vehicles, the project can be used to control speed of autonomous cars. GPS technology does not suffer from the aforementioned setbacks. The GPS receiver makes use of the satellites and the infrastructure already in place and an extensive library of map information is available. Therefore, the project uses GPS receivers to alert the driver about excess speeds at a given location. The project highlights how the information can be used to communicate with the on-board controller of a given vehicle and also how the speed of its vehicle is controlled. Several output peripherals such as LCD display and LEDs are used to provide feedback to the driver.

#### **IV) PROPOSED SYSTEM**

“IOT Based Vehicle Speed Control by Using Mobile” is developed. When driver increase the speed of vehicle then according to our algorithm if speed of vehicle is greater than average speed, then application user will receive the alert message through the cloud server as shown in working diagram, and user have a an authority to control the speed of vehicle according to his/her choices remotely at any location. Using this we can avoid the accidents and provide the safety of drivers as well as public. Safety is a necessary part of man’s life. Due to the accident cases reported daily on the major roads in all parts of the developed and developing countries, more attention is needed for research in the designing an efficient car driving aiding system. It is expected that if such a device is designed and incorporated into our cars as a road safety device, it will reduce the incidence of accidents on our roads and various premises, with subsequent reduction in loss of life and property.

#### **V) WORKING OF SYSTEM**

Following figure shows the Working of project. It shows an Interaction between the vehicle and application (User) through the middleware (cloud server).

When speed is above than the average speed then, firstly vehicle sensors send the alert to cloud server after that cloud server will forward the alert message to mobile user (admin). After receiving alert user have can change the speed of vehicle by select the appropriate speed form the application, so that vehicle speed will be limit after selection. This will be done by sending request to the server some instruction to cloud and cloud forward this instruction vehicles Driver. if driver reduce the speed itself then process is break otherwise authority to reduce the speed of vehical and set the limit of fixed speed range of speed.





### **FUTURE SCOPE**

- we can modify the system with the help of Tracker to identify the zones
- we can also modify the system with efficient braking system in association with air flow control in the carburetor.
- this system can be more efficiently use in any automobile such as lorries, buses, car.
- In future This control system is fully worked on eBike.

### **REFERENCES**

- [1.]Khan, M. A., & Khan, S. F. (2018, April). IoTbased framework for Vehicle Over-speed detection. In 2018 1st International Conference on Computer Applications & Information Security (ICCAIS)
- [2].Pérez, J., Seco, F., Milanés, V., Jiménez, A., Díaz, J. C., & De Pedro, T. (2010). An RFID-based intelligent vehicle speed controller using active traffic signals.Sensors, 10(6), 5872-5887.
- [3].K. N. V. Satyanarayana\*, G. Yaswanthini, P. L. Kartheeka, N. Rajkumar, A. BhimaRaju. (2018). IOT Based Vehicle Speed Control Automatically in Restricted Areas using RFID.IJET,7(3.31), 72-74
- [4].John, A., &Nishanth, P. R. (2017, April). Real time embedded system for accident prevention. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 2, pp. 645-648).IEEE.
- [5].Bhaskar,Seelam. Ch. Vijaya and S, Anitha, Improvement of QoS in Wireless Sensor Networks by Minimizing Path Delays (2018). International Journal of Advanced Studies of Scientific Research, Vol. 4, No. 1, 2019

## Word Sense Disambiguation for Marathi Language in Cross Language Information Retrieval

Vivek A. Manwar<sup>1</sup>, Rita L. Gupta<sup>2</sup>, Dr. A. B. Manwar<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Sant Gadge Baba Amravati University  
[vivek.manwar007@gmail.com](mailto:vivek.manwar007@gmail.com)

<sup>2</sup>Research Scholar, Department of Computer Science, Sant Gadge Baba Amravati University  
[rita25gupta@gmail.com](mailto:rita25gupta@gmail.com)

<sup>3</sup>Associate professor, Department of Computer Science Sant Gadge Baba Amravati University

### ABSTRACT: -

In natural language processing some words have multiple senses (meanings). Due to multiple sense's problems creates ambiguity in sentences. This challenge accepted by Word Sense Disambiguation. It is one of the challenges in NLP (Natural Language Processing) and it occurs in all the languages. Human can easily disambiguate the word but the machine cannot. Word Sense Disambiguation (WSD) is the task of perfectly assigning the correct sense (meaning) to the words having multiple senses in the given natural language text. The work carried out on Marathi language is very less.

### KEYWORDS: -

Information Retrieval, Cross-Language Information Retrieval, Word Sense Disambiguation, Multi-lingual Information Retrieval, Cross-lingual Information Retrieval, Natural Language Processing

### INTRODUCTION

Information retrieval is the science of retrieving information relevant to information seekers from the collection of information resources such as text, images, documents, audio, as well as music. The increasing necessity of multilingual documents Retrieval in response to the user query opens up a new branch of Information Retrieval called as Cross-Language Information Retrieval. [1]. One of the key problems in IR is related to the multiple representation of a meaning. A document is retrieved for query may be different even though the term which is occurred in query and document may same. This makes difficult to match the result of relevant document against a query. This representation problem is even more evident in cross-language information retrieval (CLIR), where queries and documents are described in different languages. Information Retrieval engines need to convert the term in various languages in which languages users enter a query [2]. The information retrieval may be classified as: Bi Lingual, Cross Lingual and Multi Lingual. Now a day, citizens of the country have more interest on global education, business and research etc. which forces them to retrieve the content from the Internet in English. But many people accessing information in a native language as they are comfortable to access the same. Cross-language information retrieval (CLIR) has a problem that is faced by certain languages where knowledge resources are limited such as Hindi, Malayalam and Marathi. India is a multilingual country. [3].

### WORD SENSE DISAMBIGUATION

Word sense disambiguation is a subtask of Natural Language Processing that deals with the problem of identifying the correct sense of the word in particular context. Indian constitution had lists of 22 languages; we referred these languages as scheduled languages. These languages are given status, recognition and official encouragement of the entire population, barely 10% Indians use English to transact and most prefer regional languages, which have evolved over centuries. India consists of the multiple states also, from all these state within the Maharashtra

region specifically Marathi is used as the regional language. But Maharashtra state again divided into multiple regions such as Vidarbha, Marathwada, South Maharashtra and Konkan. The Marathi language spoken in this entire region also differs. So, finding the sense of the Marathi word is quite complicated. Marathi language consists of multiple words that are spelled same but meaning-wise/ sense-wise they are different. Such type of words when need to be from translated from source language to target Creates the problem of Ambiguity. 1. माझी पाठ दुखत आहे (My Back is aching) 2. त्याला घडा पाठ आहे (He has *learnt* Lesson by heart) due to this problem there is a need to use Word sense disambiguation technique to find the exact sense of the word[4].

### LITERATURE REVIEW

In this section, a work done by the research community in this area in context to word sense disambiguation is presented.

**Sreelakshmi Gopal et al.** [6] implemented a Supervised Malayalam word sense disambiguation system using Naive Bayes classifier. Word Sense Disambiguation is a difficult problem in NLP because a one word may have different meanings in different situations. For all human beings it is very easy to discover the accurate sense in a particular context but for machines it is very difficult to predict. Some extents of intelligence are added to the machine for an accurate prediction.

**Krishnanjan Bhattacharjee et al.** [7] put forward gap analysis in surveyed WSD systems comparing strengths and weaknesses of various surveyed systems and their accuracy. Based on the findings, a future hybrid approach synergizing rule-based and machine learning based methods are template. All major natural languages of the world have an intrinsic semantic feature called polysemy; same word has multiple meanings as per contexts.

**Alok Ranjan Pal et al.** [8] proposed a model that disambiguates the actual sense of an ambiguous word in a particular context using Naive Bayes probability distribution. Authors have implemented their work using the Naive Bayes probabilistic model.

**Lokesh Nandanwar** [10] proposed the graph-based unsupervised Word Sense Disambiguation Algorithm to resolve the ambiguity of a word in a given HINDI Language sentence. Finding the proper meaning of a word here implies identification of the most important node from the set of graph nodes which are representative the senses. They make use of HINDI WordNet developed at IIT Bombay as reference library of words to form the sense graph. Graph based approaches in Natural Language Processing (NLP) involves the selection of best suitable candidates (node) from many interrelated candidates. In this method, graphs correspond to senses and edges corresponds to sense relations. Authors have used the HINDI WordNet to carry out the word sense disambiguation task in NLP.

**Gauri Dhopavkar et al.** [11] described a rule-based method used for performing Word Sense Disambiguation task of Text in Marathi Language. This paper states that their method successfully identifies the correct sense of the given text from the predefined possible senses using word rules and sentence rules. The system presented works on only single sentence and identifies the ambiguity. The system accuracy is around 75% which include disambiguation of nouns, adjectives and verbs in Marathi language. Authors have stated that the system can only identify and resolve word level ambiguity.

**Nutan B. Zungre et al.** [12] proposed a Graph-based algorithm, through which word ambiguity is resolved based on their senses and context domain. In Graph-based algorithm, a graph is formed that encompasses the word which is to be disambiguated with their corresponding candidate sense. In the proposed work, Marathi WordNet prepared by IIT-Bombay, multiple senses of Marathi word have been explored. The proposed system is used to figure out the rightful sense of word in Marathi language using Decision graph algorithm. The proposed system uses Source Language as Marathi. Input Text in Marathi is obtained through Google Input Tool. Marathi WordNet is used to obtain the exact features of each word in the

sentence. Marathi language is the Target language for which the sense disambiguation has been executed.

### Approaches for word sense disambiguation

Following approaches are used in word sense disambiguation.

### Dictionary-based or Knowledge-based Methods

As the name recommends, for disambiguation, these methods mostly rely on dictionaries, treasures and lexical knowledge base. They don't use corpora evidences for disambiguation. The Lesk method introduced by Michael Lesk in 1986 is the seminal dictionary-based method. The Lesk algorithm is based on Lesk definition, "measure overlap between sense definitions for all words in context". But, in 2000, Kilgarriff and Rosensweig presented the simplified Lesk definition as "measure overlap between sense definitions of word and current context", which identify the correct sense for one word at a time.

### Supervised Methods

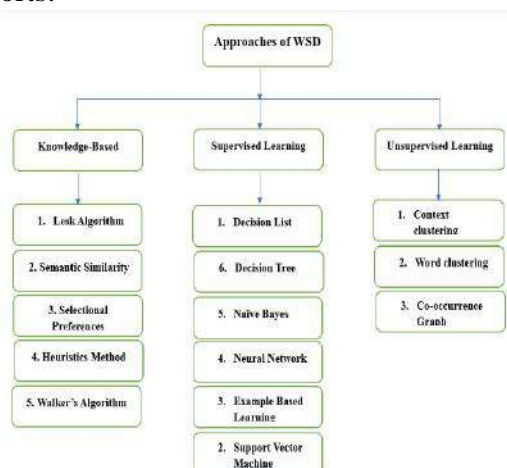
These are machine learning methods make use of sense-annotated corpora to train for the purpose of disambiguation. These methods assume that to disambiguate the sense the context can provide enough evidence on its own. In these methods, the words knowledge and reasoning are considered unnecessary. The context is represented as a set of "features" of the words it also includes the information about the surrounding words. The most successful supervised learning approaches to WSD are Support vector machine and memory-based learning.

### Semi-supervised Methods

Due to the lack of training corpus most of the word sense disambiguation algorithms use semi-supervised learning methods. Hence semi-supervised methods use both labeled as well as unlabeled data. These methods need very small amount of annotated text and large amount of plain un-annotated text. The technique which is used by semi-supervised methods is bootstrapping from seed data.

### Unsupervised Methods

Unsupervised methods assume that similar senses occur in similar context that is why the senses can be induced from text by clustering word occurrences by using some measure of similarity of the context called word sense induction or discrimination. Unsupervised methods have excessive potential to overcome the knowledge acquisition bottleneck due to non-dependency on manual efforts.



## CHALLENGES IN WSD

### Differences between dictionaries

In some cases, many senses are closely related to each other so the division of words into senses becomes much more difficult.

### Part-of-speech tagging

Part-of-speech tagging and sense tagging are closely related to each other. So it becomes difficult to decide whether these tasks should be kept together or decoupled.

### Human Interaction

Human interaction delivers the better result than computer to find the sense of words. But it is difficult for human to memorize all the possible senses of words.

### Common sense

Sometimes it is difficult to analyze the meaning of words without common sense. For example, "Ram and Shyam are fathers" and "Ram and Shyam are brothers". In first sentence each is independently a father whereas in second sentence they are brothers of each other.

### Sense Inventory and algorithms task-dependency

It is difficult to find the correct sense of word between multiple languages. For example, the ambiguity of "mouse" (animal or device) is not relevant in English-French machine translation.[15]

## GAP ANALYSIS AND CONCLUSION

The material which is available on web is mostly in English language. Hence, while retrieving the information in the regional language the user does not get the desired relevant information. The reason for that is language barriers and lack of resources available for that language hence there is need of Cross Language Information Retrieval.

Secondly, in CLIR when user enter a specific query for retrieving the data this query needs to be translated in English and then user gets the result. Most of the time relevancy of the retrieved document hampers due to unavailability of the regional words in a query from the online dictionary and also these words may have multiple meanings.

## REFERENCES: -

- [1] HL Shashirekha and Ibrahim Gashaw, "Enhanced Amharic-Arabic Cross-Language Information Retrieval System using Part of Speech Tagging", IEEE (2019).
- [2] Nurul Amelina Nasharuddin et al, "A Review on Building Bilingual Comparable Corpora for Resource-limited Languages", IEEE, Fourth International Conference on Information Retrieval and Knowledge Management, (2018).
- [3] Jay Patel et al, "Cross-lingual Information Retrieval: application and Challenges for Indian Languages", 5th International Conference for Convergence in Technology (I2CT) Pune, India (2019).
- [4] Gauri Dhovavkar et al, "Application of Rule Based Approach to Word Sense Disambiguation of Marathi Language Text", IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and communication systems (ICIIECS) (2015).
- [5] Sreelakshmi Gopal and Rosna P Haroon, "Malayalam Word Sense Disambiguation using Naïve Bayes Classifier", IEEE, International Conference on Advances in Human Machine Interaction, (2016).
- [7] Krishnanjan Bhattacharjee et al, "Survey and Gap Analysis of Word Sense Disambiguation approaches on Unstructured Texts", Proceedings of the International Conference on Electronics and Sustainable Communication Systems IEEE, (2020).
- [8] Alok Ranjan Pal and Diganta Saha, "Word Sense Disambiguation in Bengali: An Auto-updated Learning Set Increases the Accuracy of the Result", Springer Information Systems Design and Intelligent Applications, Advances in Intelligent Systems and Computing, (2016) pp. 423-430.
- [9] VARINDER PAL SINGH, "Word sense disambiguation for Punjabi language using deep learning techniques", Springer, (2019).
- [10] Lokesh Nandanwar, "Graph Connectivity for Unsupervised Word Sense Disambiguation for HINDI Language", IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication systems (ICIIECS), (2015).
- [11] Gauri Dhovavkar, "Syntactic Analyzer using Morphological Process for a Given Text in Natural Language for Sense Disambiguation", IEEE, (2014).
- [12] Nutan B. Zungre, Gauri M. Dhovavkar, "Sense Disambiguation for Marathi Language Words Using Decision Graph Method", IEEE Sponsored World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR), (2016).
- [13] Sudha Bhingardive and Pushpak Bhattacharyya, "Word Sense Disambiguation Using Indo WordNet", Springer, (2017), pp. 243-260.
- [14] Ganesh Chandra and Sanjay K.Dwivedi, "A Literature Survey on Various Approaches of Word Sense Disambiguation", 2nd International Symposium on computational and Business Intelligence, 2014.

## Machine Learning Applications: A Comprehensive Overview of Techniques and Working Mechanisms

**Ms. Vaishnavi S. Karale**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
vaishnavikarale3921@gmail.com

**Dr. Priyanka C. Tikekar**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
priyanka.tikekar058@gmail.com

### Abstract-

This research paper delves into the multifaceted landscape of machine learning, exploring its introduction, diverse applications, techniques, and underlying working mechanisms. The introductory section provides a contextual framework, elucidating the evolution and significance of machine learning in contemporary technology. The paper then navigates through an extensive array of applications across various domains, showcasing how machine learning has revolutionized fields such as healthcare, finance, image recognition, natural language processing, and more. In the subsequent sections, the research paper meticulously elucidates prominent machine learning techniques, ranging from supervised learning to unsupervised learning, reinforcement learning, and deep learning. Furthermore, the paper elucidates the working mechanisms that underpin machine learning algorithms. A comprehensive exploration of data preprocessing, feature engineering, model training, and evaluation methodologies is presented.

**Keywords-** Machine learning, Techniques & its applications

### I. INTRODUCTION

In the era of unprecedented data generation and computational capability, the field of machine learning has emerged as a transformative force, fundamentally altering the way to approach problems and make decisions. Machine learning, a subset of artificial intelligence, empowers systems to automatically learn and improve from experience, without explicit programming. This dynamic technology has permeated diverse sectors, ranging from healthcare and finance to entertainment and beyond, revolutionizing the landscape of innovation.

At its core, machine learning involves the development of algorithms and models that enable computers to recognize patterns, make predictions, and autonomously adapt to changing circumstances. The explosive growth of data availability, coupled with advancements in computing power, has catalyzed the rapid evolution of machine learning methodologies, unlocking new possibilities in fields where traditional approaches fall short. It becomes apparent that the impact of machine learning extends far beyond mere automation. It holds the potential to drive advancements in fields such as healthcare diagnostics, financial forecasting, natural language processing, image recognition, and more, making it a cornerstone of the fourth industrial revolution.

The journey into machine learning encompasses a diverse array of techniques, including supervised learning, unsupervised learning, reinforcement learning, and the revolutionary advancements in deep learning. Each method brings its unique strengths and applications, contributing to the versatility of machine learning as a problem-solving tool. In this ever-evolving landscape, understanding the foundations of machine learning is not only a technical necessity but also a key to unlocking innovation and addressing complex challenges. As we embark on this exploration, the objective is to unravel the essence of machine learning,

unveiling its potential to transform industries, enhance decision-making processes, and pave the way for a future where intelligent systems seamlessly integrate into our daily lives [1][2].

## II. METHODS OF MACHINE LEARNING

Machine learning encompasses a variety of methods that enable systems to learn from data and make predictions or decisions without explicit programming. Here are some fundamental methods of machine learning [3][4]

- A. *Supervised Learning*
- B. *Unsupervised Learning*
- C. *Reinforcement Learning*

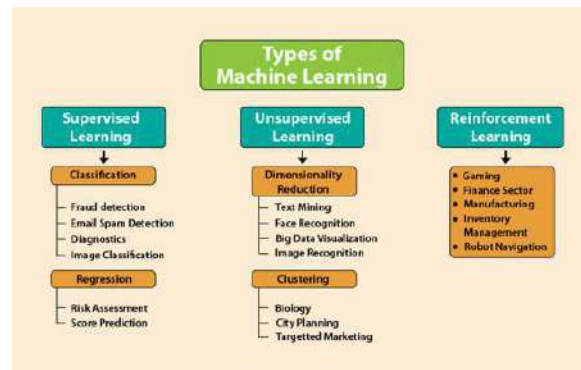


Fig. 1 Types of machine learning

### A. *Supervised Learning*

Supervised Learning is a method where both input and output data are provided to the computer during the training process, accompanied by feedback. The computer's predictions are evaluated for accuracy during training. The primary objective of this training is to enable computers to learn the mapping between input data and corresponding output. In essence, the system learns to generalize from the provided examples, making accurate predictions for new, unseen data based on the learned patterns and relationships [5][6].

### B. *Unsupervised Learning*

Unsupervised Learning operates without explicit training, where computers are not provided with input-output pairs. Instead, they autonomously discover patterns and relationships within the data. This approach is commonly applied to transactional data and is particularly useful for more complex tasks. Unsupervised learning often employs an iterative process, such as deep learning, to derive meaningful conclusions from the data. By exploring the inherent structures and associations within the input data, the system gains insights without the need for labeled examples, making it well-suited for tasks where explicit guidance is challenging or impractical. [7].

### C. *Reinforcement Learning*

Reinforcement Learning is characterized by its reliance on three key components: the agent, the environment, and actions. The agent is responsible for perceiving its surroundings, while the environment is the setting in which the agent interacts and takes actions. The fundamental objective of reinforcement learning is to determine the optimal policy, guiding the agent in making decisions and taking actions within the given environment. Through a continuous process of interaction and feedback, the agent learns to identify the most effective strategies or policies that lead to favorable outcomes, aligning with the overarching goal of achieving optimal performance in a dynamic and evolving environment [8][9].

## III. WORKING OF MACHINE LEARNING

---

The working of machine learning involves a series of steps that enable systems to learn from data, make predictions, and improve performance over time. Here is a simplified overview of the key components and processes involved in the working of machine learning [10]

*A. Data Collection:*

The first step is to gather relevant and representative data that the machine learning system will learn from. This data may include input features and corresponding output labels in the case of supervised learning.

*B. Data Preprocessing:*

Raw data is often noisy or incomplete. Preprocessing involves cleaning, transforming, and organizing the data to ensure it is suitable for training the machine learning model. This may include handling missing values, normalizing features, or encoding categorical variables.

*C. Model Selection:*

Choosing an appropriate machine learning model is crucial. The selection depends on the nature of the problem (classification, regression, clustering) and the characteristics of the data. Common models include decision trees, support vector machines, neural networks, and more.

*D. Training the Model:*

During the training phase, the selected model is fed with the prepared data. In supervised learning, the model learns to map input features to output labels by adjusting its parameters. In unsupervised learning, the model identifies patterns and structures within the data.

*E. Evaluation:*

The trained model is then evaluated on a separate set of data not used during the training phase (testing set or validation set). This allows assessing the model's performance and generalization to new, unseen data.

*F. Hyperparameter Tuning:*

Fine-tuning the hyperparameters of the model is crucial to improve its performance. Hyperparameters are settings that are not learned from the data but influence the learning process, such as the learning rate or the depth of a decision tree.

*G. Prediction and Inference:*

Once the model is trained and evaluated, it can be used to make predictions on new, unseen data. The model applies the learned patterns to generate predictions or classifications.

*H. Feedback Loop and Iteration:*

In many cases, the machine learning system operates in an iterative manner. The model's predictions are compared to the actual outcomes, and this feedback is used to update the model further, enhancing its accuracy and effectiveness over time.

*I. Deployment:*

After successful training and evaluation, the model is deployed for practical use. It can be integrated into applications, systems, or processes to automate decision-making based on the learned patterns.

*J. Monitoring and Maintenance:*

Continuous monitoring of the model's performance is essential. If the data distribution changes or the model's effectiveness degrades over time, retraining or updating the model may be necessary.

Understanding the working of machine learning involves grasping these fundamental steps, from data collection to model deployment, and appreciating the iterative nature of refining models based on feedback and new data [11].



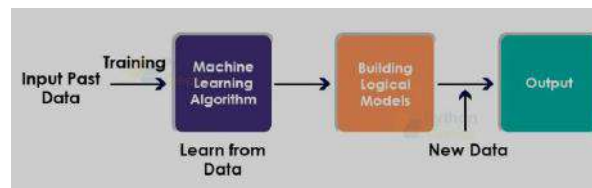


Fig. 2 Working of machine learning

#### IV. MACHINE LEARNING APPLICATIONS

Following are some of the applications of machine learning:

A. *Cognitive Services:*

Machine learning is employed in cognitive services to enhance capabilities like natural language understanding, speech recognition, and computer vision, enabling more intuitive interactions between humans and machines.

B. *Medical Services:*

In the medical field, machine learning finds applications in disease diagnosis, personalized treatment plans, predicting patient outcomes, and drug discovery, contributing to advancements in healthcare.

C. *Language Processing:*

Language processing applications leverage machine learning for tasks such as language translation, sentiment analysis, and speech synthesis, making communication more efficient and accessible.

D. *Business Management:*

Machine learning aids in business management through applications like demand forecasting, customer relationship management, fraud detection, and optimization of supply chain processes.

E. *Image Recognition:*

Image recognition systems utilize machine learning to identify and categorize objects within images. This technology is widely applied in areas like security, autonomous vehicles, and quality control.

F. *Face Detection:*

Face detection algorithms, a subset of computer vision, employ machine learning to locate and identify faces in images or videos. Applications include security systems, social media tagging, and photo organization.

G. *Video Games:*

Machine learning is utilized in the gaming industry for creating realistic and adaptive gameplay experiences. This includes non-player character behavior, dynamic game environments, and personalized gaming experiences.

H. *Computer Vision:*

Computer vision, powered by machine learning, enables machines to interpret and understand visual information from the world, leading to applications such as object recognition, image segmentation, and scene understanding.

I. *Pattern Recognition:*

Machine learning algorithms excel in recognizing patterns within data, leading to applications in fields such as finance, manufacturing, and cybersecurity, where identifying anomalies or trends is crucial.

These applications illustrate the diverse impact of machine learning across various domains, showcasing its ability to automate, optimize, and innovate processes in numerous fields [12][13].

## V. CONCLUSION

In conclusion, this research paper has provided a comprehensive exploration of the multifaceted realm of machine learning, encompassing its introduction, applications, methods, and working principles. The introduction of machine learning elucidated the fundamental concepts of it, elucidating its role as a transformative force in adapting systems to learn from experience without explicit programming. Subsequently, a journey through diverse applications showcased the ubiquitous influence of machine learning across industries, from healthcare and business management to language processing and image recognition. The myriad of methods, ranging from supervised and unsupervised learning to reinforcement learning and deep learning, illustrated the versatility of machine learning methodologies, each uniquely suited to address specific challenges. The fusion of innovative applications, diverse methodologies, and a deeper comprehension of working principles positions machine learning as a cornerstone in solving complex challenges and driving progress across various domains.

## REFERENCES

- [1] S. Angra and S. Ahuja, "Machine learning and its applications: A review," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, Andhra Pradesh, India, pp. 57-60, doi: 10.1109/ICBDACI.2017.8070809, 2017.
- [2] H. Wang, C. Ma and L. Zhou, "A Brief Review of Machine Learning and Its Application," International Conference on Information Engineering and Computer Science, Wuhan, China, pp. 1-4, doi: 10.1109/ICIECS.2009.5362936, 2009.
- [3] P. Louridas and C. Ebert, "Machine Learning," in IEEE Software, vol. 33, no. 5, pp. 110-115, Sept.-Oct. doi: 10.1109/MS.2016.114, 2016.
- [4] H. M. E. Misilmani and T. Naous, "Machine Learning in Antenna Design: An Overview on Machine Learning Concept and Algorithms," International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, pp. 600-607, doi: 10.1109/HPCS48598.2019.9188224, 2019.
- [5] S. Singh, K. R. Ramkumar and A. Kukkar, "Machine Learning Techniques and Implementation of Different ML Algorithms," 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, pp. 1-6, doi: 10.1109/GCAT52182.2021.9586806, 2021.
- [6] V. Gupta, V. K. Mishra, P. Singhal and A. Kumar, "An Overview of Supervised Machine Learning Algorithm," 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, pp. 87-92, doi: 10.1109/SMART55829.2022.10047618, 2022.
- [7] A. Srivastava, S. Saini and D. Gupta, "Comparison of Various Machine Learning Techniques and Its Uses in Different Fields," 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, pp. 81-86, doi: 10.1109/ICECA.2019.8822068, 2019.
- [8] D. Kataria et al., "Artificial Intelligence And Machine Learning," 2022 IEEE Future Networks World Forum (FNWF), Montreal, QC, Canada, pp. 1-70, doi: 10.1109/FNWF55208.2022.00133, 2022.
- [9] O. Obulesu, M. Mahendra and M. ThirilokReddy, "Machine Learning Techniques and Tools: A Survey," International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 605-611, doi: 10.1109/ICIRCA.2018.8597302, 2018.
- [10] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, Thailand, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629, 2017.
- [11] Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. Electron Markets 31, 685–695 <https://doi.org/10.1007/s12525-021-00475-2>, 2021.
- [12] Y. Kumar, K. Kaur and G. Singh, "Machine Learning Aspects and its Applications Towards Different Research Areas," International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, pp. 150-156, doi: 10.1109/ICCAKM46823.2020.9051502, 2020.  
T. R. N and R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques," IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, pp. 1-6, doi: 10.1109/SCEECS48394.2020.190, 2020.

## Intricacies of VR Features in graphic designing

**Dr. Bhargavi S. Chinchmalatpure**

Assistant Professor  
Bharatiya Mahavidyalaya, Amravati  
bhargavicm16@bmv.ac.in

**Mr. Pratik Vilas Dabherao**

Msc 1<sup>st</sup> year 1<sup>st</sup> SEM  
Bharatiya Mahavidyalaya, Amravati  
pratikdhaberao@gmail.com

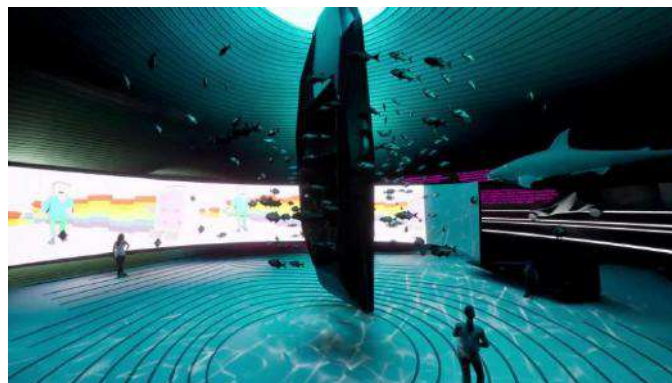
### Abstract: -

The arrival of the information age has promoted the rapid development of information technology, and the birth of VR technology has brought many changes to the emerging technology fields. This research paper explores advanced techniques in computer graphics, focusing on real-time rendering and interactive visualizations it incorporates elements of virtual reality, offering an immersive experience through the integration of VR technology. The outcomes showcase a blend of artistic creativity and technical proficiency, pushing the boundaries of visual representation in the digital sphere. Virtual reality technology has evolved for decades since its birth. As a new technology, VR has greatly changed people's daily life. No matter in the field of graphic design or interior design, it can be used as a special technical means to effectively penetrate between abstract thinking and entity, thus promoting the change of people's design thinking. Therefore, VR technology has been paid more and more attention in recent years, so it is widely used. In this paper analyses the application of VR technology in graphic design in detail, sorts out the advantages and existing problems of VR technology in graphic design field.

**Key Words:** - Real-time rendering, 3D modelling, Computer-generated imagery (CGI), Virtual reality (VR)

### 1. INTRODUCTION

In the ever-evolving landscape of computer graphics, this research paper seeks to explore and implement advanced techniques to enhance visual representation in digital environments with a focus on real-time rendering, shader programming, and interactive visualizations. Incorporating elements of virtual reality, we delve into the immersive potential of VR technology. Through a fusion of artistic creativity and technical innovation this endeavour aims to contribute to the cutting edge of computer graphics, offering a glimpse into the exciting possibilities that emerge at the intersection of art and technology figure 1 shows architecture of virtual environment [1].



**Fig. 1. The Architecture of Virtual Environment**

---

## 2.VR Technology

Virtual Reality (VR) technology has emerged as a transformative force in the realm of computer graphics. By immersing users in simulated environments, VR transcends traditional boundaries of visual interaction. This research paper delves into the intricacies of VR, exploring its integration with computer graphics to create immersive and interactive experiences. Through this convergence of graphics and virtual reality, we strive to contribute to the ongoing evolution of interactive and immersive computing environments [2].

Virtual reality (VR) technology includes simulation technology, computer plane technology, man-machine interface technology, multimedia technology, sensing technology, network technology, and other technologies [3]. It will generate a simulation environment through the above technologies and conduct a systematic simulation. The system simulates multi-information fusion, interactive 3D dynamic vision, and entity behaviour, and users can immerse themselves in the virtual environment, which is the uniqueness of VR technology.

## 3.VR Feature

A standout feature VR is the seamless integration of Virtual Reality (VR) technology. Through the utilization of VR, users are transported into immersive digital realms, experiencing interactive and lifelike environments. The research paper delves into the intricacies of VR features, including stereoscopic visuals, precise head tracking, and spatial audio [4]. This not only enhances the user experience but also opens avenues for novel applications in fields ranging from entertainment to education. The research paper's emphasis on VR as a central feature reflects a commitment to pushing the boundaries of computer graphics and creating engaging, realistic virtual experiences.

## 4.Technologies in VR:

Virtual environment refers to the use of computer plane system and various interface devices to display and control the interactive three-dimensional environment generated on the computer.

VR technology is an important direction of simulation technology. simulation technology and computer graphics, man-machine interface technology, multimedia technology, sensor technology and network technology, and other cutting-edge technology, and the combination of as a cross technology and VR technology has certain difficulty; at the same time, VR technology is a very challenging frontier discipline and research area.[5]

Virtual Reality (VR) is noted for its exceptional capability to induce a profound state of immersion, a sensation crafted by the complex interplay of its technological elements, reproducing a believable replica of reality within the user's sensory landscape environment. The technology fabricates lifelike visuals and audio, giving birth to a thoroughly immersive experience with transformative potential [6]. Four technologies are crucial for VR.

- The visual (and aural and haptic) displays that immerse the user in the virtual world and that block out contradictory sensory impressions from the real world.
- The graphics rendering system that generates, at 20 to 30 frames per second, the ever-changing images. The tracking system that continually reports the position and orientation of the user's head and limbs.
- The database construction and maintenance system for building and maintaining detailed and realistic models of the virtual world [6].

■ synthesized sound, displayed to the ears, including directional sound and simulated sound fields. display of synthesized forces and other haptic sensations to the kinaesthetic senses.

■ Devices, such as tracked gloves with push buttons, by which the user specifies interactions with virtual objects, interaction techniques that substitute for the real interactions possible with the physical world. [7]

### 5. Displays Technologies in VR:

Display technology has advanced very rapidly, pulled along by the television, presentation-research paperion, and LCD-device markets, rather than just the still-small VR market. As VR was developing, much ink was spilled over the relative merits of various formats of displays (fig.2) head-mounted displays (HMDs), CAVE-like (Cave Automatic Virtual Environment), and desktop displays. Most workers consider desktop displays not to be VR because they hardly block out the real world, do not present virtual-world objects in life size, and therefore do not create the illusion of immersion. [8]



**Fig.2 Model engineering in VR World:**

Now that we can explore quite large virtual world models in real time, we find that acquiring, cleaning, updating, and versioning even static world models is itself a substantial engineering task. It resembles software engineering in magnitude and in some, but not all, other aspects.

#### 5.1 Tracking sensors :

Tracking sensors are essential to allow the user to roam in real environment, in order to move their viewpoint in AR/VR, and be able to continuously update the location of the user in the virtual world. For this reason, it is regarded as one of the main components of VR/AR systems. These sensors basically interact with a system processing unit, relaying the orientation of the user's point of view to the system [9]. Using the sensors in combination with VR systems does not only allow the detection of the location of the user, it also helps detect the user's direction of movement and speed of that movement in any direction. The race is on between the different VR hardware and software developers and technology for VR is so rapidly evolving that reviewing the latest technology specifications.

**5.2 Stereoscopic imagery:** A binocular HMD can display slightly different viewing angles for each eye, creating binocular overlap, which gives the viewer the illusion of stereoscopic depth and a more or less realistic 3D viewing experience, creating the illusion that some objects are near and others far away[10]. It is important that the binocular overlap is correct, because if it is not, the user will experience an unfocused, double-vision image.

**5.3 Interpupillary distance (IPD):** The distance between the two eyes, measured at the pupils. It is important that a head-mounted display has an adjustable IPD, because the effectiveness of the stereoscopic imagery depends on it and every end-user has a slightly different IPD setting at which they will experience a clear focused image.

**5.4 Field of view (FOV):** The natural FOV of human beings is about 180°, but so far HMDs are not capable of creating this. Current HMDs offer a field of view of between 60° to 150°. A bigger field of view results in a more realistic user experience, with a greater sense of immersion, allowing the user to establish greater situational awareness and more effective interaction[11].

### **5.5 VR in hospitality & tourism**

VR application areas for Hospitality and Tourism are still under development as the technology becomes more mainstream, both the industry and the consumers are starting to appreciate the possibilities this technology has for their hospitality and tourism interests. As the technology matures, the application areas are rapidly being explored by the early-adopters. This section analyses the immediate application areas for the Hospitality and Tourism industry and highlights the opportunities that lie ahead as the technology evolves. The implementations of VR technologies continue to impress consumers and investors and as a result, these increasingly sophisticated technologies are being envisioned and implemented for end-user benefits in the Tourism and Hospitality industry.

## **6. Importance of VR Technology**

The following analysis highlights the impact and importance of VR technology in graphic designing:

**6.1 Effective Planning and Suitable Management:** With the help of VR technologies in the potential has widened in terms of implementing effective policy and also effective planning. VR devices create almost realistic, easy and detailed navigation.

**6.2 Effective Entertainment Tool:** Considering the important history of VR devices starting with the introduction of the "Sensorama Simulator" in 1962, which provided people a virtual experience of driving a motorcycle, including realistic movement, sound, scent and airflow, VR technology has evolved to much greater heights since then[12].

**6.3 Education Tool:** VR has tremendous potential in terms of education and effective research of many years has proved that VR devices, technology can serve as a great tool for entertainment. A VR model can be an efficient means of communication of large amounts of information because it leverages the user's natural spatial perception abilities [13]. VR has great potential to entertain and educate people via games, interactive sessions, Artificial Intelligence based Interactive Systems and many more

## **7. Conclusion:**

This research paper has journeyed through the dynamic landscape of computer graphics, exploring advanced techniques that redefine visual representation in the digital realm. As we reflect on the accomplishments achieved, the research paper stands as a testament to the continuous evolution of computer graphics and its profound impact on diverse fields, promising exciting avenues for future exploration and innovation. It is believed that virtual reality technology will have more and more applications in the field of graphic design in the future. There may be many problems in the application of virtual reality technology in the field of graphic design, so it needs the efforts of people from all walks of life and support for graphic

design. In the future, VR technology will bring huge economic benefits to related industries, and at the same time, they will lead the future of related industries.

### References :

1. Junxiang wang , siru chen , yuxuan liu , and richen lau, "Intelligent Metaverse Scene Content Construction", vol. 11, pg No.76222- 76241, July 2023.
2. R. Liu, M. Gao, L. Wang, X. Wang, Y. Xiang, A. Zhang, J. Xia, Y. Chen, and S. Chen, "Interactive extended reality techniques in information visualization," IEEE Trans. Hum.-Mach. Syst., vol. 52, no. 6, pp. 1338–1351, Dec. 2022.
3. R. Wang, L. Jiang, J. Yuan, X. Xu, and W. Wang, "Virtual reality scene construction based on multimodal video scene segmentation algorithm," in Proc. IEEE, pp. 1817–1820, May 2019
4. Lulu Liu, Minh Tien Nhung, "The Application of VR/AR Technology in Graphic Design Based on zSpace", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1668296, 7 pages, 2022.
5. Hale KS & Stanney KH, "Handbook of Virtual Environments: Design, Implementation, and Applications", Second Edition, CRC Press, (2014).
6. Schmalstieg D & Hollerer T, Augmented Reality: Principles and Practice, Addison-Wesley, (2016).
7. K. Kroes, I. Saccardi, and J. Masthoff, "Empathizing with virtual agents: The effect of personification and general empathic tendencies," in Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality (AIVR), Dec. 2022, pp. 73–81.
8. Goh C, Mok H & Law R, "Artificial intelligence applications in tourism in Khosrow-Pour", Vol.2., pp. 22–25 Aug2018.
9. Mueen uddin , selvakumar manickam , hidayat ullah , muath obaidat , and abdulhalim dandoush, "Unveiling the Metaverse: Exploring Emerging Trends, Multifaceted Perspectives, and Future Challenges", vol 11, 87087 IEEE 2023.
10. I.-C. Stanica, F. Moldoveanu, G.-P. Portelli, M.-I. Dascalu, A. Moldoveanu, and M. G. Ristea, "Flexible virtual reality system for neurorehabilitation and quality of life improvement," Sensors, vol. 20, no. 21, p. 6045, Oct. 2020.
11. F. Dudyrev, O. Maksimenkova, and D. Mikhailenko, "Intelligent virtual reality tutoring systems as a new generation of simulators: Requirements and opportunities," in Proc. IEEE Global Eng. Educ. Conf. (EDUCON), pp. 706–718 Apr. 2021.
12. Eduard Petlenkov, Aleksei Tepljakov, and Kristina Vassiljeva "Virtual Reality Meets Intelligence in Large Scale Architecture Ahmet Kose(B)", Springer International Publishing, pp. 297–309, 2017.
13. Man, W., Qun, Z. "The deconstruction and reshaping of space: the application of virtual reality in living space", (IEEE), pp. 410–413. January 2017.

## Review Paper on Characteristics, Benefits and Challenges in Cloud Computing

**Miss. Sidhdee Satish Gurjar.**

M.Sc.II<sup>nd</sup> Year,  
Department of Computer Science,  
Vidya Bharati Mahavidyalaya,  
Amravati  
Email ID: [sidhdeegurjar23@gmail.com](mailto:sidhdeegurjar23@gmail.com)

**Miss. Adika B. Thakare.**

M.Sc.II<sup>nd</sup> Year,  
Department of Computer Science,  
Vidya Bharati Mahavidyalaya,  
Amravati  
Email ID: [adikathakare2001@gmail.com](mailto:adikathakare2001@gmail.com)

**Prof. D. M. Kene**

Professor,  
Department of Computer Science,  
Vidya Bharati Mahavidyalaya,  
Amravati  
Email ID: [dkene75@gmail.com](mailto:dkene75@gmail.com)

### ABSTRACT: -

Cloud computing is a rapidly emerging technology that has significantly changed and opened up potential for many Indian industries. This ubiquitous computing paradigm has completely changed the way that information technology services and infrastructure can be provided. The use of cloud computing in the educational field is becoming more popular. The goal of the current study is to give a general overview of the cloud computing concept and its uses for student and academic collaboration. We suggested cloud computing for e-learning in this study, taking into account its advantages, disadvantages, work mode, services, and benefits. This essay examines cloud computing's potential educational applications, with a focus on management schools. A primary study was conducted with key players in the technical education infrastructures that are put into place for academic usage. The most recent research on cloud computing's application in education was carried out using a qualitative approach. Approximately eight research papers have been discovered and presented following a thorough study of the literature to demonstrate the significance and likely applications of cloud computing in the field of education. The primary stakeholders in education are identified, and the benefits and hazards associated with using cloud computing are analysed in this survey. According to a thorough examination, it is possible to offer more clarity on the benefits of cloud computing by integrating it into management education.

**KEYWORDS: -** Cloud Computing, Infrastructure-as-a-service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), Data storage as service (DaaS), Public Clouds, Private Clouds, Hybrid Clouds, Community Clouds.

### INTRODUCTION: -

When it comes to computing resources, cloud computing refers to on-demand network access that is typically offered by a third party and only needs minimal supervision. Networks, servers, storage, apps, and services are some of these resources. When combined with other technologies and design techniques, cloud computing offers a variety of designs and realistic models.

One of the top ten disruptive technologies for the upcoming years, according to Gartner, is cloud computing. It symbolizes the long-cherished goal of seeing computing as a service, wherein the principles of economies of scale significantly reduce the cost of computer infrastructure. Leading companies, including Sun Microsystems, Google, IBM, Amazon, and Microsoft, have started building new data centers for hosting cloud computing applications in various parts of the world in order to ensure consistency and provide redundancy in the event of a site failure or collapse.

Because all data and apps are stored online, cloud computing offers a number of benefits along with several drawbacks. Due to the real-time and online accessibility of cloud-based apps and data, it can be applied to a variety of daily tasks, such as teaching. A concept known as "cloud computing" makes it possible to have easy, on-demand network access to a



shared pool of resources (such as servers, storage, apps, and services), which can be quickly supplied and released with little administration labor. According to the adopted cloud model, its use forms the foundation of its monetizable worth.

Both instructors and students can access a wide range of cloud-based apps and services for both formal and informal education. Increased collaboration, communication, and resource sharing are made possible by cloud computing, which also offers more mobility and flexibility in the usage of resources for teaching and learning. It also generates virtual communities of teaching and learning or individualized learning environments.

### **HISTORY OF CLOUD COMPUTING: -**

Cloud computing was developed by John McCarthy in 1960. "The use of computers as a subject of research may be arranged as a public utility eventually." According to Parkhill in The Computer Utility Challenges [1], the name "cloud" computing was introduced in the telecommunications industry as a virtual private network. . There was a waste of bandwidth using point-point data lines. Network utilization was balanced using a virtual private network. Servers and network infrastructure are now included. Cloud computing has been widely used by industry participants. Amazon introduced Amazon Web Services, and this has been of great help to their business. Furthermore, Google and IBM have both launched cloud computing research. Eucalyptus was the first open-source platform for private cloud deployment.

### **CLOUD COMPUTING ARCHITECTURE: -**

There are three categories into which cloud computing services fall. the ends on the front and rear. The network, which is typically the network, connects the front end and back end[1]. A system's client, or user, sees its front end, while the system cloud is its back end. Many applications, such as web-based bulk processing systems for the back office, use cloud architectures. Here are a few examples.

- Processing pipelines for OCR-based document processing: This creates searchable raw text from millions of pages and images and converts thousands of Microsoft Word documents to PDF.
- Image processing pipelines capable of encrypting AVI or MPEG files. Constructing a web crawler index Data mining is used to sift through millions of documents.
- Batch processing systems are a kind of back-office program that are used in the retail, banking, and insurance sectors. Reports are produced on a daily and weekly basis using log analysis.

#### **i. Infrastructure-as-a-service (IaaS):**

Users of the cloud directly utilize the processing, storage, networks, and other computer resources and IT infrastructure that are made available by the cloud. Virtualization is widely employed in the (IaaS) cloud to mix and match physical resources as needed to satisfy the fluctuating resource requirements of cloud customers. The fundamental method of virtualization is building unique virtual machines (VMs) that are separated from one another and from the underlying hardware. The multitenancy paradigm modifies the software architecture of the program by enabling several instances (from different cloud users) to function on a single application. This strategy is different from that model. A few instances of infrastructure as a service include Google, App Engine, Microsoft Azure, Java, and developer tools[4].

#### **ii. Platform as a service (PaaS):**

Cloud users can create cloud services and applications by using a platform called "platform as a service," which supports the entire "software lifecycle." This

offers a development platform that hosts both completed and unfinished cloud applications, in contrast to SaaS, which only hosts completed cloud applications. Therefore, in addition to a hosting environment, PaaS provides development infrastructure, including tools, programming environments, configuration management, and other components. Examples of Platform as a Service (PaaS) include Java, Microsoft Azure, Google App Engine, and developer tools.

iii. **Software as a service (SaaS):**

Programs published in a hosting environment by cloud clients are accessible to a wide range of users with network access (such as web browsers). The SaaS cloud groups users of different cloud consumers' applications into a single logical environment to optimize speed, availability, disaster recovery, maintenance, and security, as well as to realize economies of scale. The cloud infrastructure is not controlled by the user and often uses a number of system architectures. Salesforce, Google Docs, and Google are a few examples.

iv. **Data storage as service (DaaS):**

Data storage services are now a distinct cloud service that offers virtualized storage that is made available on demand. A great data storage solution is available as a special kind of IaaS. This is because dedicated servers, software licenses, post-delivery services, and internal IT maintenance can sometimes come with costly upfront costs for on-premises enterprise database systems. Instead of getting a site license for the whole database, customers can utilize DaaS to pay only for the services they use. Some data storage service providers offer table-style abstractions that store and retrieve large amounts of data in very short times, in addition to more conventional storage interfaces like file systems and relational database management systems, which are often too big, too slow, and quite expensive.

### **CHARACTERISTICS OF CLOUD COMPUTING: -**

The National Institute of Standards and Technology[2] lists five characteristics of cloud computing—such as resource pooling, broad network access, rapid elasticity, etc.—that make it appropriate for use in information technology services and applications.

i. **On-demand self-service:**

Cloud services can be automatically given to customers as needed, without the need for human interaction. These services include server time, storage, web applications, processing power, and networks.

ii. **Resource pooling:**

To serve numerous clients, cloud services combine their computing resources. Either "multi-tenancy," which lets several users share resources, or virtualization, which uses virtual computers to imitate physical hardware, are used to achieve this.[5]

### **TYPES OF CLOUDS: -**

There are three types of cloud computing: private, hybrid, and public clouds[6].

- i. **Public Clouds:** Companies that use and control public clouds do so to provide other organizations and individuals with quick and reasonably priced access to computer resources. It is not necessary for consumers to buy hardware, software, or auxiliary infrastructure when using public cloud services because these are owned and maintained by the providers. A service provider that hosts the cloud infrastructure makes public clouds accessible to everyone. A few instances of public clouds are

- Google AppEngine, Sun Cloud, IBM's Blue Cloud, Amazon Elastic Compute Cloud (EC2), and Windows Azure Services Platform.
- ii. **Private Clouds:** Private clouds are data center architectures with provisioning, automation, monitoring, scalability, and flexibility that are owned by a specific company. The purpose of a private cloud is to obtain the advantages of cloud architecture without giving up control over your own data center maintenance, as opposed to selling "as-a-service" solutions to outside clients. Compared to public clouds, private clouds are more expensive but also more secure.
  - iii. **Hybrid Clouds:** A hybrid cloud is made up of two or more clouds—public, private, or community—that continue to exist as separate entities but are connected to provide the benefits of various deployment options. You can have complete or partial control over third-party cloud providers in a hybrid cloud, which increases computing flexibility. For example, specific apps or sections of applications can be moved to the public cloud during busy times.
  - iv. **Community Clouds:** The purpose of a community cloud is to serve its needs. These communities are made up of individuals or groups with similar interests. This covers groups for standards, industry, research, and so forth. A hybrid type of private cloud designed and run especially for a particular group is called a community cloud. These communities aim to collaborate in order to accomplish their shared business goals, and they have comparable cloud requirements. These clouds aim to provide participating enterprises with the extra privacy, security, and policy compliance typically associated with a private cloud while still enabling them to enjoy the benefits of a public cloud.

#### **BENIFITS OF CLOUD COMPUTING: -**

Users are encouraged to adopt cloud computing because of its many benefits. Using cloud computing has several advantages, the main ones being easy scalability, cost savings, and greater productivity.

- i. **Cost reduction:** By using software as a service, businesses can reduce the amount they pay for IT resources, which boosts their operations' productivity and profitability. Payments must be made by customers based on their usage.
- ii. **Increase productivity:** As a result of the quick development of technology, customers are expecting more from brands. Online or cloud-based computer systems must be able to access business applications for cloud computing. Availability of the programmers, which are always and everywhere available to consumers.
- iii. **Scalability:** On-demand business scalability is made possible by the scalable concept of cloud computing. SaaS, PaaS, and IaaS are a few examples. A company can always employ fewer virtual servers than it currently needs based on service demand. There is no set price that small businesses must pay for hosting in a dedicated data centre.

#### **CHALLENGES OF CLOUD COMPUTING: -**

There are many obstacles associated with cloud computing technology for various data and information handling sectors. Thus, in the event that you decide to implement cloud infrastructure services, you can run into the following challenges and risks[3]:

- i. **Security and privacy:** These include the technological and organisational challenges associated with preserving an adequate degree of data security and privacy in cloud services. This ensures that when government organisations use the cloud, major security and privacy issues pertaining to the security and privacy of sensitive or important data for a business, like banks, will surface.

- ii. **Data Management:** The demand for effective data management solutions has grown as a result of the greater number of data-intensive applications that cloud computing enables on the widest possible scale. Additional issues with cloud computing across several data centres include data processing and retrieval.
- iii. **Service Management:** The cloud-based IT approach presented a number of challenges for service management. Another challenge is the ability to provide more context-sensitive and customised services.

**CONCLUSION: -**

The architecture, types, and characteristics of cloud computing were covered in this study. Cloud computing is important to information technology since it lowers costs for businesses and facilitates file access. It also aids in lowering redundancy and data latency. The two main obstacles to cloud computing adoption for any firm are security and privacy.

**REFERENCE: -**

- [1] S. A. Sheik and A. P. Muniyandi, "Secure authentication schemes in cloud computing with glimpse of Artificial Neural Networks: A Review," *Cyber Security and Applications*, vol. 1, p. 100002, 2023.
- [2] P. mell and T. Grance, "The NIST Definition of cloud computing Recommendations of standards and Technology" Nat.l Stand. Technol.lab, Vol. 145, p.7, 2011.
- [3] VeritisAdmin, "Cloud computing trends, Challenges & Benefits," *Go to Veritis Group Inc.*, 12-Oct-2022. [Online]. Available:
- [4] "12 benefits of cloud computing and its advantages," *Salesforce.com*.
- [5] P. M. Mell and T. Grance, "The NIST definition of cloud computing," 2011.
- [6] <http://www.appcore.com/types-cloud-computing-private-public-hybridclouds/>

## A Review on Big Data Challenges and Hadoop Technology

**Kavita Kishor Yadav**

Asst. Prof. Dept. of Comp Sci.  
Vidya Bharati Mahavidyalaya, Amravati.  
Kavitakyadav.30@gmail.com

### Abstract

Big Data refers to the huge volume of data which is being processed on a daily basis and needs to be stored, distributed and managed. It can be structured, semi-structured or unstructured. Parallelism is used to process this data in an efficient manner. Big Data demands a platform and a technique which can analyse this data and extract hidden knowledge from it. Hadoop is an open source software, a core platform to structure Big Data. It offers a distributed file system known as HDFS (Hadoop Distributed File System) and MapReduce structure to deal with large data and provide a high degree of fault tolerance. This paper aimed at study and analysing big data concepts and Hadoop technology. This study designed about big data concepts, its issues and challenges. This paper also consists of Hadoop concepts and it's framework. This paper mainly explains the Big Data, its issues and challenges. Along with this an introduction to Hadoop and its components and characteristics is also in this paper.

**Keywords :** Big Data, Types of Data, Issues and Challenges, Hadoop, Hadoop Framework

### Introduction

#### Big Data

In digital world, data are generated from various sources and the fast transition from digital technologies has led to growth of big data. It provides evolutionary breakthroughs in many fields with collection of large datasets. In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management tools or data processing applications.[1]

This proliferation of data is primarily driven by advancements in technology, the widespread use of the Internet, social media platforms, mobile devices, and the Internet of Things (IoT). The data generated from these sources include structured, semi-structured, and unstructured data, creating a massive volume and variety of information.



Fig.1: Big Data

## Types of Big Data

**1 Structured Data:** Structured data refers to organized and well-defined data that fits into a fixed schema or format. It can be easily stored, managed, and processed using traditional relational databases. Structured data includes information such as numbers, dates, names, addresses, and categorical variables. Ex. customer transaction records, inventory databases, and financial statements.

**2 Semi-structured Data:** Semi-structured data possesses some organization but does not conform to a rigid schema. It contains metadata or tags that provide limited structure and enables better organization and search ability. Ex. XML files, JSON documents, and log files. This type of data is often encountered in web applications, social media feeds, and data exchanged between different systems.

**3 Unstructured Data:** Unstructured data refers to data that lacks a predefined structure and does not fit neatly into traditional databases. It includes textual data, multimedia content, social media posts, emails, sensor data, and more. Unstructured data is often challenging to analyse due to its complexity, large size, and lack of organization.

**4 Quasi-structured Data:** The data format consists of unstructured textual data that lacks consistent formatting and requires significant effort and time to organize using specialized tools. For instance, web server logs are log files generated and maintained by servers, which contain a list of activities and events recorded on the server. These logs often include information such as IP addresses, timestamps, HTTP requests, and response codes[5].

## Issues and Challenges

Challenges in big data can be broadly alienated into three types the first type is data challenges, the second type is data process challenges, and the third type are data management. Data challenges are the challenges that are associated with the characteristics of big data. Process challenges are the challenged that faced during the processing of data whereas management challenges pertaining to tackling the data such as providing security.

The characteristics of big data bring many challenges to it such as its high volume, variety, etc. Process challenges are related to data acquisition, pre-processing, data analysis, and data visualization whereas management challenges are related to privacy and security[4].

1. Data Challenges Researchers have given many definitions of big data and based on their understanding towards they come up with several new characteristics of big data. researchers discussed the 3V's characteristics of data (Volume, Variety, and Volume), 4th V was introduced by IBM as veracity, researchers have discussed 5th and 6th V's as variability, and value. The 10 V's of big data are taken under consideration, there are many worth mentioning prominent challenges associated with the characteristics of data. Some of the prominent challenges are discussed as follow.

**1.1. Volume Challenges.** The unprecedented increase in data through internal and external sources has resulted in a massive amount of data. This high volume of data brings the challenges to the data itself such as the storage of the data for processing is not possible through traditional tools and thus more innovative methods should be developed to handle this data deluge.

**1.2. Variety Challenges.** The challenge associated with variety is related to its different forms. The massive data can be present in the form of structured, semi-structured, and unstructured.

---

Research studies show that 95% of the data is present in unstructured form. Therefore, converting it into a form so that the analysis can be performed is a big challenge.

**1.3. Velocity Challenges.** Velocity indicates the speed of the data generated through the devices. Data can be processed in two ways batch processing and real-time processing. In batch processing, the data is stored and then processed whereas real-time processing is continuous. In online shopping, real-time processing is required to generate value for customers.

**1.4. Veracity Challenges.** Data veracity indicates the quality and accuracy of data. It deals with the fabrications, imprecision, messiness, and misplaced evidence in data. It defines the trustworthiness of data when a significant decision needs to be taken. In social networking sites, user opinion can be classified as positive, negative, or neutral.

**1.5. Value Challenges.** Value is one of the most significant features of big data characteristics. Big data contains valuable information that needs to be extracted from the large datasets. This brings a big challenge to data as extracting the high information from data in a cost-effective manner and making use of it for business intelligence, health sectors, etc.

**2. Process Challenges** Process challenges are related to processing and analyzing large datasets. It brings a significant challenge to the process as the data is present in different forms and conversion of it into one form for analysis purpose is a challenging task. It can be divided into four parts: Data Acquisition and Storage, Data Preprocessing, Data Analysis and Modeling, Data Visualization.

**2.1. Data Acquisition and Storage.** Data acquisition is the process of acquiring and storing the data for the future utilizing some valuable information. The data is acquired from various sources such as from sensors, social networking sites, blogs, etc. and hence the data is present in different forms (structured, semi-structured, and unstructured) bring a significant challenge to data. The second challenge is associated with the storage because the data generated through various devices does not mean that whole data carry meaning to it therefore the smart filter must be applied for generating the relevant datasets. Storing this massive dataset can result in high-cost scalable systems to handle the data.

**2.2. Data Preprocessing.** Data preprocessing is the process to collect the quality data from large datasets as low-quality data leads to low-quality knowledge. Therefore, data preprocessing plays a significant role in knowledge discovery. In this stage noise, missing values, inconsistent and superfluous data, etc. are removed before applying the big data mining techniques to its data. In big data preprocessing most of the efforts are done in the Feature Selection method whereas some of the families of it are ignored such for instance reduction, missing value imputations, noise treatments.

**2.3. Data Analysis and Modelling.** Data analysis is the process that discovers the hidden information from the data and helps the organizations to make a better decision. To efficiently extract the knowledge from the large datasets extraordinary techniques are required. To generate the hidden pattern from the large datasets Wal-Mart's employs statistical and machine learning techniques.

**2.4. Data Visualization.** A big data visualization technique presents the analytical data visual form. It makes usage of various types of graph for representing the valuable information for decision making. As per the research studies, the visual report has a better impact on information seeker rather than the text reports. Visualization tools like Tableau and Qlik View are the tools used for visualization however according to the researcher these tools cannot be fruitful shortly where data is growing every second by each one of us.

**3. Management Challenges** Management challenges are related to those challenges encountered by an organization which is related to the privacy, security, governance of data. Management challenges are also faced because we have a lack of data for skilled professionals who know the latest tool and techniques to employ the correct method for dealing with each phase of data. Security and privacy will always be the major concerns as data are highly sensitive such as financial data, military data, insurance codes and contains different kinds of information that can ruin if the unauthorized user has access to it [4].

## **Hadoop**

Hadoop has become a very popular platform in the IT industry and academia for its ability to handle large amounts of data, along with extensive processing and analysis facilities. Different users produce these large datasets, and most of data are unstructured, increasing the requirements for memory and I/O. Besides, the advent of many new applications and technologies brought much larger volumes of complex data, including social media, e.g., Facebook, Twitter, YouTube, online shopping, machine data, system data, and browsing history. This massive amount of digital data becomes a challenging task for the management to store, process, and analyse. The conventional database management tools are unable to handle this type of data. Big data technologies, tools, and procedures allowed organizations to capture process speedily, and analyse large quantities of data and extract appropriate information at a reasonable cost. Several solutions are available to handle this problems. Distributed computing is one possible solution considered as the most efficient and fault-tolerant method for companies to store and process massive amounts of data. Among this new group of tools, Map Reduce and Spark are the most commonly used cluster computing tools. They provide users with various functions using simple application programming interfaces (API)[9].

## **Hadoop Framework**

Hadoop is open any one software used to process the Big Data. It is very famous used by administrations/researchers to analyze the Big Data. Hadoop is influenced by Google's structural design, Google File System and MapReduce. Hadoop procedures the large data sets in a spread calculating environment.

Hadoop contains of two main mechanisms:

### **1) Storing:**

The(HDFS) Hadoop Distributed File System: These are dispersed file system which brings responsibility taking and measured to run on creation hardware. HDFS brings high amount entree to application data and is suitable for requests that have vast data sets. HDFS can stock data over thousands of servers. HDFS has master/slave construction . Files added to HDFS are separated into fixed-size masses. Mass size is configurable, but avoidances to 64 megabytes. [7]



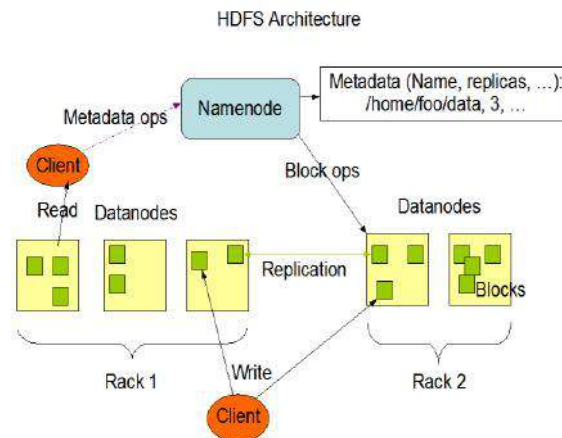


Fig-2: HDFS Blocks

## 2) Processing:

**MapReduce** : It is a software project classical presented by Google in 2004 for effortlessly writing applications which procedures enormous volume of data in equivalent on huge bunches of hardware in responsibility. This functions on huge data set, separations the problem and data sets and run it in equivalent way. Two utilities in MapReduce are as following:

a) **Map** –The Map function continually runs first naturally used to filter, transform, or parse the data. The outcome from Map develops the input to Reduce.

b) **Reduce** –The Reduce function is elective normally used to encapsulate data from the Mapfunction. [2]

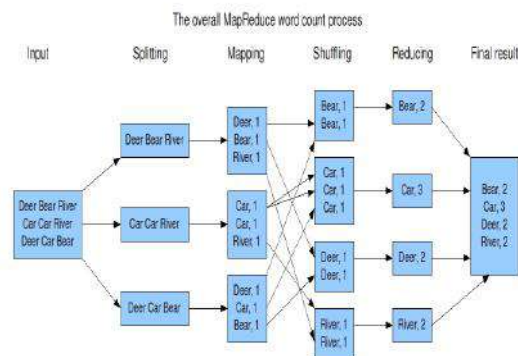


Fig-3: Map Reduce Processing

## Hadoop Characteristics in Big Data

1. **Robust**: It can handle failures of hardware of data is stored in multiple data nodes.
2. **Scalable**: it can able to increase cluster size by add more and more nodes.
3. **Simple**: It focuses on code rather than data and is can write parallel wise so it's simple.
4. **Portable**: Because Structured (In table format), Semi structure- Not in well organized format (XML), Unstructured. It is no format (Text, Image, Videos).
5. **Cost Effective**: Hadoop uses commodity hardware to store the data so it is inexpensive and economic.

6. Fault Tolerance: the tasks are automatically redirected to another node if a node fails it is fault tolerance automatically stored multiple copies of all the data. If one node fails, same data is available on some other nodes is based on replication factor.[6]

### **Hadoop for Big Data**

Effective storage, computation, and analysis of large volumes of data are major challenges of big data. Earlier due to the less advanced technology unstructured data is not handling by several organizations. So Hadoop big data changed that way and decision-making process is being used for unstructured data. Hadoop provides a reliable and scalable platform which is used to solve problems caused by massive amount of data. Hadoop big data is popular because of the properties like flexibility, scalability, performance, and cost effective. MapReduce is a programming framework which is used to processing and analyzing the big data in a cost-effective manner. Hadoop data analytics ecosystem includes data storage, data processing, data access, data management, privacy data protection[10].

### **Conclusion**

Nowadays everyone has many data which is unstructured and collected from various sources. All this data is big data. In this review paper, we described the overview of Big data and Hadoop technology. We discussed the various problems of Big data and then we discussed about its solution (Hadoop). The paper describes Hadoop which is an open source software used for processing of Big Data, along with its features HDFS and MapReduce.

### **References**

- [1] D. P. Acharjya, Kauser Ahmed P, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016
- [2] Rahul Beakta, Big Data And Hadoop: A Review Paper, <https://www.researchgate.net/publication/281403776>
- [3] Nikhil Madaan, Umang Kumar, Suman Kr Jha, Big Data Analytics: A Literature Review Paper, International Journal OF Engineering Research & Technology (Ijert)
- [4] R Rawat and R Yadav, Big Data: Big Data Analysis, Issues and Challenges and Technologies
- [5] 1Mohammad Nazmul Alam, 2Vakil Singh, 3Ms. Ripendeep Kaur, 4Md. Shahin Kabir, Big Data: An Overview with Legal Aspects and Future Prospects, Journal of Emerging Technologies and Innovative Research (JETIR) [www.jetir.org\(ISSN-2349-5162\)](http://www.jetir.org(ISSN-2349-5162))
- [6] N. Deshai1 , S. Venkataramana2 , Dr. G. P. Saradhi Varma, Research Paper on Big Data and Hadoop-Map Reduce Real Time Scheduling, International Journal of Advance Research in Science and Engineering [www.ijarse.com\(ISSN: 2319-8354\)](http://www.ijarse.com(ISSN: 2319-8354))
- [7] Vinayak Pujari1, Dr. Yogesh K. Sharma2 and Rohan Rane, A REVIEW PAPER ON BIG DATA AND HADOOP, International Journal of Advance and Innovative Research(ISSN: 2394-7780)
- [8] Pankaj Saraswat1 , and Swapnil Raj, A Review Paper on Hadoop Architecture, International Journal of Innovative Research in Computer Science & Technology (IJRCST),(ISSN: 2347-5552) [www.ijrcst.org](http://www.ijrcst.org)
- [9] N. Ahmed , Andre L. C. Barczak , Teo Susnjak and Mohammed A. Rashid / A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench / Ahmed et al. J Big Data (2020) 7:110 <https://doi.org/10.1186/s40537-020-00388-5>
- [10] Toshifa, Aniruddh Sanga, Shweta Mongia/ Big Data Hadoop Tools and Technologies: A Review/ International Conference on Advancements in Computing & Management (ICACM-2019)
- [11] [WWW.GoogleScholar.com](http://WWW.GoogleScholar.com)

## A Review on Internet of Things (IoT) Sensors

**Mithilesh M. Wasu**

Research Scholar,  
Dept. of Electronics and Computer Science,  
PGTD, RTM, Nagpur University,  
Nagpur, MS, India

mithileshwasu546@gmail.com

**Dr. Kishor M. Dhole**

Assistant Professor,  
Dept. of Computer Science,  
Seth Kesarimal Porwal College of Arts and  
Science and Commerce, Kamptee, Nagpur,  
MS, India

km.phd108@gmail.com

**Abstract:** In recent years, Internet of Things (IoT) is one of the most rapidly growing and major contributing technology for improving the quality of life. It is indeed the future of communication that has transformed real world objects (Things) into smarter devices. This paper provides an overview on IoT, its characteristics, and summarizes the classification of various sensors and actuators.

**Keywords:** Internet of Things (IoT), Characteristics of IoT, Sensors and Actuators, Types of Sensors.

### Introduction

The 21<sup>st</sup> century era is marked with a fast-paced world where all things are interconnected with each other. The IoT is one of the major contributing technologies of this time comprising of a huge network of connected things that gather, store, and use data about their surrounding environment. The ‘Things’ in the IoT context are nothing but the devices connected of daily use like ovens, room heaters, smartphones, lights, TVs, wearable fitness devices, etc. Hence, an IoT is called an ecosystem of the interconnected machines [1].

The term “Internet of Things” (IoT) was coined by Kevin Ashton in 1999. He is one of the founders of the Massachusetts Institute of Technology’s Automatic Recognition Lab. Ashton initiated RFID technology in the field of Supply Chain Management (SCM). The IoT is an emerging topic of technical, social and economic importance. The IoT spans its breadths to consumer products, goods, automobiles, industrial components and facilities, sensors, and other everyday objects combined with internet connectivity and powerful data analysis capabilities.

The “Thing” in IoT can be any device with any type of sensor embedded with the ability to collect data and transmit it across the network without manual intervention. The technology embedded in the object helps to interact with internal states and the external environment, which in turn aids in the decision-making process.

IoT is comprised of various areas including cloud, mobile devices, virtualized environments, sensors, Radio Frequency Identification (RFID), and Artificial Intelligence. In any IoT application, sensors bring the physical world very close to the digital world that can be implemented by leveraging fog computing.

### IoT Definitions

The IEEE defines IoT as “The Internet of Things aims at offering new applications and services bridging the physical and virtual worlds, in which Machine-to-Machine (M2M) communications represents the baseline communication that enables the interactions between Things and applications in the cloud”.

Whereas the Oxford dictionaries describes IoT as “Internet of Things (noun): The interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data”. The IoT can be thought as an inclusive information system comprising of smaller information systems [2].

### Characteristics of IoT

The IoT can be described using the following features.

1. **Interconnectivity:** With regard to the IoT, the global information and communication infrastructure can interconnect anything.
2. **Things-related services:** The IoT is capable of providing things-related services. To deliver object-related services within the constraints of things, both the physical world technologies and the information world will be changing.
3. **Heterogeneity:** The IoT devices are heterogeneous, based on different platforms and networks of hardware. They are capable to interact through different networks with other devices or service platforms.
4. **Dynamic Changes:** Device status changes dynamically, e.g. sleeping and waking up, connecting and/or disconnecting, as well as device context including location and speed. In addition, the number of devices can dynamically change.
5. **Connectivity:** Connectivity allows accessibility and compatibility of the network. Accessibility becomes available on a network and the compatibility provides the ordinary capability to consume and generate data.
6. **Sensor:** A sensor is generally a device capable of detecting changes in an environment. A sensor is able to measure a physical phenomenon and transform it into an electric signal [3].

### IoT Actuators and Sensors

The sensors and actuators are the foundational blocks for building any IoT application. In many IoT applications, you need one or more sensors to collect data and information about the system. The data is then collected, transmitted over the network and the actuators that allow things to work. Sensors and actuators enable IoT solutions in every IoT verticals, from smart cities to smart farming, and from personal health to smart logistics and transportation.

#### Actuators:

Actuators are mechanical or electro-mechanical devices that provide controlled, and sometimes limited movements or positioning that are actuated electrically, manually or by various fluids such as air, hydraulic, etc. The actuators are broadly classified as:

- Linear actuators that convert power into linear motions, usually for positioning applications (Electric and Hydraulic).
- A hydraulic actuator consists of a cylinder or fluid drive that uses hydraulic power to facilitate mechanical operation.
- Electric actuators may provide operating power/torque in one of several ways.
- Electromechanical actuators can be used to drive a motor that converts electrical energy into mechanical torque.
- Rotary actuators convert energy to provide rotational motion.
- Pneumatic actuators use compressed air that allow large forces to be produced from relatively small pressure changes. A pneumatic actuator converts the energy formed by the vacuum or compressed air at high pressure into linear or rotational motion [4].

#### Sensors:

Sensors play an important part in the automation of any application by measuring and processing the collected data for detecting changes in physical things. Whenever there is a

change in any physical condition for which a sensor is made, it produces a measurable response. There are different types of sensors which can range from very simple to complex. The classification of sensors can be based on their specifications, its conversion method, type of material used, its sensing physical phenomenon, properties that what it measures, and the application field.

A sensor is defined as a device to detect changes in an environment. Sensor is used to measure physical quantity (light, heat, motion, moisture, pressure, temperature) and convert them into electrical pulses. Sensor has certain properties such as sensitivity, resolution and range.

### Types of Sensors

- A. **Bio-Sensor:** Bio-sensor is a chemical sensor subset, but is often treated as a separate area. A biosensor is a self-contained analytical device that selectively and reversibly responds to the concentration or activity of biological sample chemical species, meaning that any sensor that is physically or chemically operated in biological samples can be considered as a biosensor using living components or a product of living things for measurement.
- B. **Chemical Sensors:** Sensors which response by sensing any chemical reaction, chemical substance or a set of chemicals is known as chemical sensor. A chemical sensor is used to measure the chemical composition of the environment. These types of sensors can be used for detecting environmental events, building health, agriculture conditions, and etc. Chemical sensors are widely used in industrial purposes for sensing a change in liquid or air and found to be useful for detecting healthy industrial environment.
- C. **Gas sensors:** They use various sensing technologies such as electrochemical, photoionization and semiconductor for detection of toxic gases.
- D. **Gyroscope Sensors:** Gyroscope sensors detect any tilt or angular movement in the object by measuring angular velocity. It is widely used in 3D mouse games, for the training of sportspersons, robotics, industrial automation and many more.
- E. **Humidity Sensor:** Humidity is water in the air. The presence of water vapor also influences various physical, chemical, and biological activities and its measurement in industries is critical because it can affect the product's quality and cost, staff's health and safety. Humidity sensing is therefore important for industrial processes and human life control system. Many industrial, agricultural and domestic applications are important for controlling or monitoring humidity.
- F. **Infrared Sensors:** Infrared sensors emit or detect infrared radiations in order to sense some characteristics of certain objects. They can also measure heat emission. Home automation for monitoring and controlling home appliances like turning on/off lights are the common application area of these sensors.
- G. **Micro Sensor:** Micro sensor is an extremely small device that can collect and relay information about the environment. Such devices can measure and send biological, thermal, chemical, and the other data forms to a processor, which then converts the information into a meaningful form for a variety of users to allow access to it.
- H. **Motion Sensors:** A motion detector is a device used to sense all the kinetic and physical movement in the environment. These are primarily used in automated systems like door controls, parking systems, sinks, toilet flushers and hand dryers.
- I. **Occupancy Sensors:** They are sometimes called as presence sensor; detects the presence of human or objects in a particular area. It can be used for remote monitoring through various parameters like temperature, humidity light, and air.

- 
- J. **Odour Detection Sensors:** The odour detection systems can be commonly arranged into four categories such as gas sensors, bio-sensors, gas chromatography systems and hybrid systems. Other common odour detection instruments like electronic noses (E-noses), mass spectrometers (MS), differential optical absorption spectrometers (DOAS) are frequently used sensors. The choice of any odour detection technique is influenced by several factors like on-field deploy ability, odour detection and classification capabilities, sensitivity of the instruments and sensitivity of the instruments to the air matrices. Among all the achievable detection methods, electronic nose is having highly developed capabilities as compare to other instruments. The presence of interference noise acts as a main interference for the efficiency of the odour detection devices.
- K. **Optical Sensors:** Optical sensors are useful in detecting the electromagnetic energies like light. Being passive to all forms of electrical interfaces, these are widely used in IoT applications like in digital cameras. Optical sensors are good for IoT applications related to energy, health, environment, oil refineries, chemical, industries, aerospace and other areas.
- L. **Physical Sensor:** The physical sensors generally measure physical quantities such as length, temperature pressure, electricity, weight, sound, etc. It can be defined as a device that corresponds to physical property, called stimulus, and produces a corresponding electrical signal that can be measured.
- M. **Proximity Sensors:** The position of any nearby object can easily be detected with proximity sensor without any physical contact. By emitting electromagnetic radiation such as infrared, it finds the presence of an object by simply looking for any variation in the return signal. There are different types of proximity sensors like Inductive, Capacitive, Ultrasonic, Photoelectric, Magnetic and etc. targeting different applications. Sensors of proximity are the best way to detect any movement. In applications such as safety or efficiency, they are widely used. These sensors are used as the best possible sensor for map building to avoid obstacles in navigating to a crowded place or any complex route. As proximity sensors sense objects in their vicinity, they find applications in retails, surveillance, metal and non-metal objects detection, etc.
- N. **Position Sensors:** The position sensor detects the presence of human or objects in a particular area by sensing their motion. It can be used in health care monitoring for monitoring the position of patients, nurses and doctors in a hospital, in agriculture for detecting the position of cattle.
- O. **Pressure Sensor:** A pressure sensor is a sensor which helps to sense the pressure and converts that into an electrical signal. The value of pressure sensor correlates to the pressure applied. These types of sensors can be used in health monitoring and weather forecasting, agriculture, smart vehicles, manufacturing, water system maintenance and aircrafts.
- P. **Smoke sensors:** Smoke sensors are commonly used in homes and industrial applications.
- Q. **Temperature Sensor:** These are one of the sensors commonly used to measure a given medium's temperature or heat. These sensors use a number of methods to determine and quantify and object's temperature. Some of the temperature sensors required physical contact with the object while other types do not require contact as they can detect liquid or gasses that emit radiant energy such as heat spike or temperature spike. Temperature sensors finds their use in manufacturing, agricultural, and health sectors.
- R. **Velocity sensors:** This sensor calculates the rate of change in constant position measurement and position values at known intervals. Velocity sensor may be linear or angular. A linear velocity sensor detects the speed of an object along a straight line

whereas angular velocity sensor detects how fast a device rotates. It can be used in smart city applications for intelligent vehicle monitoring.

- S. **Water Quality Sensors:** Water quality sensors are used to measure by water quality. Commonly measured parameters include water temperature, PH, turbidity, conductivity, dissolved oxygen, etc [1] - [6].

## Conclusion

This paper presented an overview of IoT, its definitions, and characteristics. Furthermore, researchers described various types of actuators and sensors. The main contribution of this paper is classification of sensors.

## References

- [1] S. Ratnaparkhi *et al.*, "WITHDRAWN: Smart agriculture sensors in IOT: A review," *Mater. Today Proc.*, p. S2214785320387447, Dec. 2020, doi: 10.1016/j.matpr.2020.11.138.
- [2] R. A. Radouan Ait Mouha, "Internet of Things (IoT)," *J. Data Anal. Inf. Process.*, vol. 09, no. 02, pp. 77–101, 2021, doi: 10.4236/jdaip.2021.92006.
- [3] J. Amalraj, S. Banumathi, and J. John, "IOT Sensors And Applications: A Survey," *Int. J. Sci. Technol. Res.*, 2019, Accessed: Jan. 30, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/IOT-Sensors-And-Applications%3A-A-Survey-Amalraj-Banumathi/dc79e0a25ac68e568fe9e8913024d1c2a35cd2c1>
- [4] D. Sehrawat and N. S. Gill, "Smart Sensors: Analysis of Different Types of IoT Sensors," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India: IEEE, Apr. 2019, pp. 523–528. doi: 10.1109/ICOEI.2019.8862778.
- [5] Manav Rachna International Institute of Research & Studies, Faridabad, India., P. Manhas, S. Thakral, Manav Rachna International Institute of Research & Studies, Faridabad, India., J. Verma, and Manav Rachna International Institute of Research & Studies, Faridabad, India., "IoT Sensors: Perspectives & Appliance," *Int. J. Recent Technol. Eng. IJRTE*, vol. 8, no. 6, pp. 3306–3310, Mar. 2020, doi: 10.35940/ijrte.F8594.038620.
- [6] R. Nandhini and S. Poovizhi, "(PDF) ARDUINO BASED SMART IRRIGATION SYSTEM USING IOT," in *ResearchGate*, Accessed: Jan. 30, 2024. [Online]. Available: [https://www.researchgate.net/publication/321854296\\_ARDUINO\\_BASED\\_SMART\\_IRRIGATION\\_SYSTEM\\_USING\\_IOT](https://www.researchgate.net/publication/321854296_ARDUINO_BASED_SMART_IRRIGATION_SYSTEM_USING_IOT)

## A Survey on Predictive Analysis of Social Media Data Using Machine Learning Algorithm

**Ms. Sheetal M. Yawalkar**

Research Scholar S.G.B.A.U.  
Amravati, MS India  
sheetalyawalkar@gmail.com

**Prachi Mawale**

Bharatiya Mahavidyalya  
Amravati, MS India  
prachimavale5@gmail.com

### **Abstract-**

Machine Learning (ML) is a branch of artificial intelligence. New technologies are constantly being developed and improved. Machine learning plays a key role in predictive analytics in many areas, including examining social media data. Social media is a rich source of information for studying human behaviour and relationships. However, the demographics and behaviours of social media users are often unknown or incomplete, limiting the applicability and effectiveness of social media analysis. This paper, provide a comprehensive survey of multiple applications of SM analysis using robust machine learning algorithms. Initially, we discuss a summary of machine learning algorithms, which are used in SM analysis. After that, we provide a detailed survey of machine learning approaches to SM analysis.

**Keywords-** *Machine learning methods, predictive analysis, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning*

### **I. Introduction**

The popularity of social media platforms is increasing, with millions of photos being published on the Internet every day [1]. In the digital world, online content has become a favourite, whether it is important information and entertainment for Internet users or businesses. For example, every minute, users post more than 300,000 tweets, more than 680,000 pieces of content on Facebook, and upload 100 hours of video to YouTube. [2] Users spend an average of 2 hours and 24 minutes on social media platforms every day. 85% of the 5.27 billion mobile phone users worldwide use social media. Table1 shows With 1.02 billion users, China is the country where social media is used the most. India and the United States round out the top three with 755.47 million and 302.25 million users respectively, and by 2024 the number of global media users will reach 9.45 billion.

As shown in fig [1]. The average social media user jumps from 6 to 7 platforms every month. Although there are many factors contributing to user growth, the global penetration of smartphones is the most obvious.

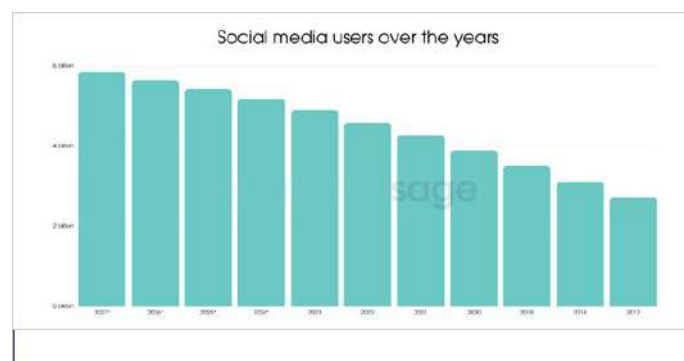


Fig [1] Social Media Users OF 10 years



Year	Social media user
2027*	5.85 billion
2026*	5.64 billion
2025*	5.42 billion
2024*	5.17 billion
2023	4.95 billion
2022	4.59 billion
2021	4.26 billion
2020	3.9 billion
2019	3.51 billion
2018	3.1 billion
2017	2.73 billion

Table 1 . Social Media Users OF 10 years (in billion)

Prediction in social media is the process of using data analysis and machine learning techniques to predict future behaviours, trends, or outcomes of users of social media, content, or platforms. Predictions can prove valuable to different aspect, online users can filter more easily the huge amount of information; content producers and content providers can better organize their information and build more effective delivery platforms; and advertising networks can

design more Sophisticated and profitable advertising strategies. However, predicting the popularity of web content, as useful as it seems, has been employed in few real-world applications. It is more beneficial to use predictive analytics on social media. Prediction is beneficial and helpful for social analysis because it helps businesses, organizations, and individuals make better decisions, improve their policies, and pursue their goals. In this research paper .we review the current practical uses of these methods and propose new applications that could benefit from this research area.

The main contributions of this paper are:

- 1) Brief study of ML Algorithm and various approaches.
- 2) Study of various ML techniques
- 3) Literature Review of research paper using machine learning algorithm.

## II. BACKGROUND

### ALGORITHMS OF ML AND APPROACHES

The word, technique will be used to refer to a type of learning (supervised, unsupervised, semi-supervised, or reinforcement) for the term, method, it will be used to refer to different methods of Machine Learning: ANN, SVM, DT, etc. We will retain the nuance between algorithm and model. A model is a set of hypotheses about a problem domain, expressed in a precise mathematical form, which is used to create a Machine Learning solution [5]. Whereas an algorithm is simply a set of instructions used to implement a model to solve a problem or perform a calculation.

In this study, we consider four types of learning: supervised learning, unsupervised learning, semi-supervised learning, and extended learning.

**Supervised Learning** can be used when historical data on the problem is available. The system is trained using common inputs and responses and is then used to predict responses to new inputs [6. Supervised Learning is divided into two types: Classification and regression [3]. Classification involves finding the relationship between non-uniform inputs and non-uniform outputs. The output is also called a group or collection. In the learning step, the training data is analysed to create a classifier and this method is used to predict the names in the classification step [3]. Regression, on the other hand, involves estimating or estimating a quantity. Regression relies on statistical techniques to establish the relationship between two or more variables [3].

Unlike supervised learning, in **unsupervised learning** there is no text (no graphics displayed). The purpose of unsupervised learning is to identify the structure of the data and extract important information from it without specifying the need [3]. There are two subtypes of unsupervised learning: clustering and dimensionality reduction. Clustering involves dividing different objects into different groups so that objects in each group are as similar as possible and objects in different groups are as different as possible [3]. The purpose of reducing file size is to convert large data into a small area without losing important information in the original data [3].

**Semi-supervised learning**, as the name suggests, is a combination of the two methods mentioned above. Semi-supervised learning is often used in situations where some events (problems) have both covariate (input) and outcome (output) values, but in most events there is only the covariate value and no information about the expected outcome [4].

**Reinforcement learning** is a specialized area of machine learning that relies on doing something based on numerical rewards to achieve a goal. The point is that in the particular world called the environment, the person acting as the so-called agent does not know what is good and what is bad, but learns by trying to do what gives the most reward.

Fig.2. shows Traditional data processing in various Learning Techniques of ML.

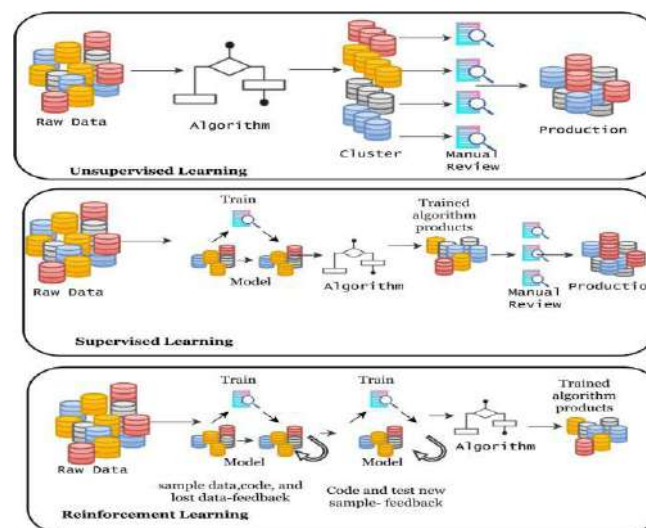


Fig.2. Traditional data processing in various Learning Techniques of ML.

### III. STUDIED MACHINE LEARNING METHODS

#### Naive Bayes

Naive Bayes is a simple classifier and can be classified quickly. Categorize "Phishing" or "Not Phishing" in the file label. For example, phishing is a popular application of the Naive Bayes algorithm [7]. Phishing filter classifier of the emails we receive to build a machine learning model, the Naive Bayes classifier uses the same set of points following Bayes' probability theorem.

#### K-means

K-means is an unsupervised regression algorithm used to anonymously group input data [8]. First, we determine the number of clusters that should be represented by K. We then choose a point K as the centroid of the group. This centroid is not provided by our data. Then, cluster K is created by assigning each data point to the nearest centroid. Calculate and place a new

centroid for each new cluster created. Finally, we re-join each data point to the nearest new centroid.

### **Support Vector Machine**

Support Vector Machine (SVM) is one of the most popular SL methods, used for regression and classification [9]. SVM classifies the given data by drawing the plane of the given data, which shows the distribution of the input data. Optimum distribution means better spacing on both sides of the plane, which is called margin maximization. Therefore, this information can be viewed in many dimensions until the results are obtained.

### **Linear Regression**

The Linear Regression (LR) algorithm [10] describes how changes in one feature will affect other features. Explains how independent properties communicate with dependent properties. Dependent attributes are the reason for the prediction; independent attribute gives a prediction meaning.

### **Logistic Regression**

Logistic Regression ML algorithm is used to classify activities. When we talk about immature traits we use logistic regression (LR)[7] and have one or more personality traits. It is similar to multiple horizontal regressions and all the same warnings apply. The goal of logistic regression is to find the best model that explains the relationship between the dual trait of interest and the process of autonomous behaviour.

### **Decision Tree**

Decision tree is a decision-based support tool frequently used in machine learning. Make a tree-like diagram where each intersection or node follows a test. The branches represent the results of these experiments. It includes time nodes, decision nodes, and end nodes. [11]In this study, we use decision trees to determine whether certain metrics contribute to prediction of certain social media popularity. This is a supervised learning algorithm.

### **Random Forest**

Random Forest, is a supervised learning algorithm [8] used for regression, classification etc. This technique is used to correct decision tree behaviour caused by over-tuning of the training process. It works by creating multiple decision trees by distributing patterns across overall dataset. It is usually done through the bagging technique. The average value of each decision tree is used as the final result.

### **Neural Networks**

Neural network models [12] are algorithms inspired by the human brain and are designed to work in the same way as our brain data. The most common neural network application is a multilayer feed forward network, which consists of multiple layers where each layer computes information from the previous layer. Inputs from one to the other are combined via a weighted linear equation. The output is modified according to the algorithm before being sent to another layer. [13]Weights are often thought of as random values passed through multiple layers. It is recommended to combine previously learned results to reduce the complexity of the calculation when solving complex problems.

## Nearest Neighbours

Nearest Neighbours also has the ability to create decision trees used in classification and regression problems [14]. The nearest neighbour classification problem is widely used in business and is a simple machine learning algorithm. This classification works as if there is a neighbour who votes for the cause, and the final classification will save that neighbour. Likewise, it accommodates all necessary issues and distributes new situations, taking into account the votes of all neighbours. We can measure this using the k nearest neighbour function for each class.

## III.LITERATURE REVIEW

Sr. No.	Study	Methods/Algorithm	Learning Techniques	Advantages	Disadvantages
1	[7]	Random Forest, KNN, AdaBoost, SVM, Naïve Bayes, Logistic Regression	Supervised and Unsupervised Learning	Random forest has performed well and given good Results.	The category feature does not provide the classification accuracy And optimization for improving popularity.
2	[8]	Random Forest ,CNN	supervised Learning	Different techniques are explored on content, contextual, and social information for improving popularity prediction.	Need to improve approach and expand multimodal dataset with additional modalities and photos.
3	[9]	Gradient boosting machine	supervised Learning	Extract numerical features in combination with text based characterization in effective way for for Social Media Headline Prediction (SMHP)	Need to be consider user based prediction based on emotion analysis.
4	[10]	Grad-CAM visualization method.	Supervised and Unsupervised Learning	This paper presented new approaches for popularity prediction of social media videos using self-attention mechanism in video frames, textual and multimodal domain.	Need to includes more fine-grained visualizations using attention on convolution feature maps, learning coordinated modality Representations, headline generation based on a video.
5	[11]	Decision Tree, Random Forests.	supervised Learning	This study gives successful analysis and prediction of Popularity of tweets with images.	Enhance parameters and image viewing as well network-related parameters may be a problem of future interest Image features and network variables did not contribute much to the popularity of the tweet.
6	[12]	CNN,GBRT,SVM, Gaussian Naive Bayes	supervised Learning	Image object detection methods were performed on images or thumbnails of videos and the results lead to adding new image related features to the dataset.	Need to add more visual features with the help of computer vision.
7	[13]	Linear Regression (LR), Random Forest regressor (RF), and Support Vector Regression (SVR)	supervised Learning	Prediction accuracy, measured with $R^2$ , can reach up to of 73.1% with all features, and 64.8% without manual image assessment.	In future research, the manual assessment values in this study can be changed to similar automated values in order to reduce subjectivity.
8	[14]	Text mining techniques	Unsupervised Learning	The paper help politicians, public Commentators, information experts, and	Data is collected from single source media and the location of the users was not

				social scientists to better understand people's views.	considered while considering English Tweets.
9	[15]	CEASE and FCA-Bass model	Supervised and Unsupervised Learning	This paper reveals an integrated model for both real-time topic detection and popularity prediction of online social network especially adopted to long-text content.	Need to be implementing on real world dataset.
10	[16]	Discrete Wavelet Transform method, Dynamic Time Warping (DTW) method, SVR, cro-SVR, TF-IDF.	supervised Learning	This model is used to predicts the event popularity based on SVR.	-

**Table 2: ML METHODS BY LEARNING TECHNIQUE, PAPER ADVANTAGES AND DISADVANTAGES**

#### IV. ANALYSIS AND DISCUSSION

In this research paper is studies current states-of-the-art on web content popularity prediction method. Presented different prediction methods .Even, if research on predicting the popularity of web content has been an active area in the largest year there are many avenues that wait to be explored. We suggest some possible direction for the future and those papers are study on methods under this category is used to predict content popularity using only the information on their web sites.

The Author P.L. Bokonda et.al.[1] in his research paper shows the most commonly used ML method are RF, ANN, DT, SVM and LR. Author Priyanka Rathord. et al. proposed[3] various classification method which evaluation parameter metrics which are computed are accuracy, precision, recall, F1, AUC. Various news outlets have been considered. Best result shown by Random Forest classification algorithms. Author Mayank Meghawat et al. proposes A Multimodal Approach to Predict Social Media Popularity [4], test results confirm that despite the author using half of the training set, the multimodal method provides comparable performance to the highest quality. A joint model using Light GBM is proposed by author Liuwu Li, et al. Author uses various data forms on the SMHP task, to improves the performance of headline prediction of social media data, textual data is used[5]. Author Adam Bielski et al. the author has proposed a way that combines Grad-CAM with a soft focus approach to visualize which video components contribute to its popularity in both local and temporary domains [6].

Author Nimish Joseph et al. proposed work [7] identifies the various parameters contribution for getting tweet viral and also help business for building of social media tools. Decision tree and random forest prediction mechanism are used for popularity prediction of tweet with images based on different factors like transactional variables, image features, network variable and historical variables of the user. Author Mehmetcan Gayberi et al. [8] predict the popularity of Instagram post, with the help of seed account specific database was created. Image acquisition methods have been developed in images or video thumbnails result in new image-related features to the database. Having a large database (210.630 posts) with multiple studies available and combining user features and post-based features based on statistical and image acquisition results in a high correlation rate of 0.92 and MAE results of approximately 0.42.

Author Kristo Radion Purba et al. performed a trend analysis and prediction on Instagram, using a set of features found in user metadata, posts, hash tags, image testing, and user history. Predictability accuracy reached 73.1% using Support Vector Regression (SVR),

Engagement Grade EG is used in comparison to respecting the low ER of users with high followers.[9]. Author Amir Karami et al. study proposes a four-step popularity analysis framework using two text mining techniques that include sentimental analysis and modeling of the topic as well with quality code[10]. Author Quanquan Chu et al. proposed [11] integrated model of both real-time discovery and predicting popularity in the context of an online social network, specially adapted to long text content. For popularity prediction author develop Feature-Combined Bass model with Association Analysis (FCA-Bass) based on time series. Author Mingding Liao et al. propose CROP, [12] cross-platform event popularity predictor model to predict event-based popularity prediction on assistive platform information.[13] Y. Sun et al. In this overview of the literature ,machine learning methods for social network analysis in social media data are examined .It looks at approaches for community recognition connection prediction ,and influence analysis ,illuminating how machine learning algorithm may extract useful information from interconnected user networks .this social media data .In order to automatically recognize and track event from user generated information and enable real-time monitoring and situational awareness ,it examines several methodologies , including supervised and unsupervised methods author D. Chicco et al.[14] Uses machine learning methods for predictive analysis of user interaction in social media data: The An-An Liu et al.[15] this study examines machine learning methods used to forecast user interaction in social media data .Szu-Chuang Li [16] et al. Highlights how predictive analysis many improve marketing may improve marketing tactics and user interaction on social media platform while examining variables including user demographics, network structure, and content characteristics

#### IV. CONCLUSION

Machine Learning for predictive analytics in social media data has enormous potential for companies and organizations looking to mine the massive amounts of data produced by social media platform for insightful information .Businesses may Identify trends, forecast results and make data driven choices by utilizing machine learning algorithms. This research paper clears idea about ML techniques, algorithm, method and models. And also various ML approaches are used to study research papers which use ML algorithm .But in order to assure the moral and responsible use of these potent technologies in the social media environment .It is crucial to negotiate the difficulties related to data privacy .Data Quality and Interpretability.

#### V. REFERENCES

- [1] P.L. Bakonda ,Nissrine souissi,Quazzani touhami Khadija “ Predictive analysis using machine learning : Review of trend and methods” conference proceedings : Nov 2020 DOI: 10.1109/ISAECT50560.2020.9523703
- [2] [Social Media Users And Statistics For 2024 \(Latest Data\) \(demandsage.com\)](https://demandsage.com)
- [3] Yang, H., Xie, X., and Kadoch, M. (2020). Machine Learning Techniques and A Case Study for Intelligent Wireless Networks. IEEE Network, 34(3), 208-215
- [4] T. Chen, E.C. Hui, J. Wu, W. Lang, X. Li, Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data, Habitat Int. 89 (2019) 102005.
- [5] Bai ,V.S., & sudha ,T.(2023).A systematic literature review on cloud forensics in cloud environment ,International Journal of Intelligent System and Application in Engineering ,11(4s),565-578.Retrieved from [www.scopus.com](http://www.scopus.com)
- [6] Dhabliya ,D. (2021). An Integrated Optimization Model for Plant Diseases Prediction with Machine Learning Model Machine Learning Application in Engineering Education and Management

- [7] Priyanka Rathord , Dr. Anurag Jain , Chetan Agrawal “A Comprehensive Review on Online News Popularity Prediction using Machine Learning Approach “ISSN 2455-0108 ,vol. 5, issue 1, January 2019, doi: <https://doi.org/10.24113/ijoscience.v5i1.181>
- [8] Mayank Meghawat ,Satyendra Yadav,Debanjan Mahata “ A Multimodal Approach to Predict Social Media Popularity” 2018 IEEE Conference on Multimedia Information Processing and Retrieval .DOI 10.1109/MIPR.2018.00042.
- [9] Liuwu Li, Sihong Huang, Ziliang He and Wenyin Liu. 2018. “An Effective Text-based Characterization Combined with Numerical Features for Social Media Headline Prediction”. In Proceedings of ACM Multimedia Conference (MM’18). Seoul, Republic of Korea.ACM. <https://doi.org/10.1145/3240508.3266438> (2018)
- [10]Adam bielski and Tomasz trzcinski “Understanding Multimodal Popularity Prediction of Social Media Videos with Self-Attention” 2018 IEEE. DOI: 10.1016/j.osnem.2019.05.002.
- [11] Nimish Joseph (&) , Amir Sultan , Arpan Kumar Kar , and P. Vigneswara Ilavarasan” Machine Learning Approach to Analyze and Predict the Popularity of Tweets with Images ©IFIP 2018 Published by Springer Nature Switzerland AG 2018. I3E 2018, LNCS 11195, pp. 567–576, 2018. [https://doi.org/10.1007/978-3-030-02131-3\\_49](https://doi.org/10.1007/978-3-030-02131-3_49).
- [12] Mehmetcan Gayberi, Sule Gunduz Oguducu “Popularity Prediction of Posts in Social Networks Based on User, Post and Image Features” ACM ISBN 978-1-4503-6238-2/19/11. doi.org/10.1145/3297662.3365812
- [13] Kristo Radion Purba, David Asirvatham, and Raja Kumar Murugesan “Instagram Post Popularity Trend Analysis and Prediction using Hashtag, Image Assessment, and User History Features” The International Arab Journal of Information Technology, Vol. 18, No. 1, January 2021 DOI :<https://doi.org/10.34028/iajit/18/1/10>
- [10] Amir Karami and Aida Elkouri “Political Popularity Analysis in Social Media” Springer Nature Switzerland AG 2019 N. G. Taylor et al. (Eds.): iConference 2019, LNCS 11420, pp. 456–465, 2019. [https://doi.org/10.1007/978-3-030-15742-5\\_44](https://doi.org/10.1007/978-3-030-15742-5_44)
- [11] Quanquan Chu ,Zhenhao Cao, Xiaofeng Gao, Peng He, Qianni Deng, and Guihai Chen “Cease with Bass: A Framework for Real-Time Topic Detection and Popularity Prediction Based on Long-Text Contents” Springer Nature Switzerland 2018 X. [https://doi.org/10.1007/978-3-030-04648-4\\_5](https://doi.org/10.1007/978-3-030-04648-4_5)
- [12] Mingding Liao1, Xiaofeng Gao, Xuezheng Peng, and Guihai Chen” CROP : An Efficient Cross-Platform Event Popularity Prediction Model for Online Media” Springer Nature Switzerland AG 2018 [https://doi.org/10.1007/978-3-319-98812-2\\_3](https://doi.org/10.1007/978-3-319-98812-2_3)
- [13]Y. Sun, C. Liang, C.-C. Chang, Online social construction of Taiwan’s rural image: Comparison between Taiwanese self-representation and Chinese perception, *Tour. Manag.* 76 (2020) 103968.
- [14] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,
- [15] An-An Liu , Xiaowen Wang,Ning Xu,Junbo Guo ,Geoqing Jin,Quan Zhang,Tang ,Shenuyan Zhang”A review of feature fusion –based media popularity prediction methods”<https://doi.org/10.1016/j.visinf.202207.003>,2022.“
- [16]Szu-Chuang Li, yu-Ching Chen ,Yi-Wen Chen, Yennun Huang” Predicting Advertisement Revenue Of Social-Media-Driven Content Websites: Toward more efficient and Sustainable Social media posting” *Sustainability* ,14(7),4225,<https://doi.org/10.3390/su>, 2022.

## **A Systematic approach for Challenges and Preventive Measures for Machine learning technology**

**Ms. Ashwini S. Kaware**

askaware7@gmail.com

Department of Computer Application (BCA)

Vidya Bharati Mahavidyalaya, Amravati.

### **Abstract:**

In recent years the cases of various cyber attacks are increasing and so there is drastic change in technology can be seen to prevent it. In today's cyber world scenario detection of these attacks is more important.

Although Machine learning is one of the most ubiquitous techniques in exposure and prevention of cyber attacks in various fields, there is also threat to such technique. In this paper, the overview of threats to such system and variety of preventive measures for each of them are provided.

**Keywords: Machine Learning (ML), Poisoning Attack, Evasion Attack, Inference Attack, Extraction Attack.**

### **Introduction:**

Machine learning is a technology that focuses on enabling computers to "learn" through "experience" instead of being directly programmed. In this the term "experience" typically means information gathered with time. Numerous benefits are offered to various organizations by automating regular tasks, forming various patterns and acquiring useful information from vast dataset. Automatically learning programs through data is more interesting than manual programming. Thus with change in time inventions and usage are increasing in field of machine learning. In the past two decades, machine learning has rapidly spread in computer science and other fields, and is broadly used in web search, spam filtering, advertising, fraud detection, and stock trading. However, all these benefits come with a threat of security. [1][2]

The poisoning, extraction, evasion, inference are some type of attack that can influence the machine learning model and can manipulate the training data set due to which the model may produce wrong results. Researchers are rapidly developing innovative defenses over such security issues to protect the integrity of system. Also the security-conscious training procedures, algorithmic enrichment and secure development practices can harden machine learning systems against common ML attacks. Technical countermeasures such as differential privacy, watermarking and model encryption of model can make strong security mechanism for ML systems [2][3].

### **Types of attacks:**

#### **1. Poisoning:**

In the poisoning attack the intruder tries to inject malicious data to ML model's training data. The main aim behind this is to force ML model to learn something it should not, which results into the inappropriate responses based on the attacker's objectives. Also the attacker can insert any harmful file to the training data of ML model which may contain malicious scripts. If such



system is used for detection of corrupted files or any viruses it could not work properly or could not recognize harmful data. Such kinds of attacks are generally found in centralized or standalone systems [4] [5].

Boiling from attack, label-flipping attack and bridging attacks are some type of poisoning attack. Among which label-flipping attack is common, in which the invader changes the label of training samples to another category but another features of data kept untouched. Due to which the ML model could not understand the sample and misclassification occurs [6].

### **Preventive Measures for Poisoning attack:**

The training data must be check and validated before using it for training machine learning model. The training data should be validated by proper security tools and there is need to make sure that data comes from trusted source. Another way of protecting ML model is anomaly detection technique on training data sets for finding suspicious samples. Use of ML models that are less vulnerable to poisoning attacks can improve the system's processing. The strength of ML model can be tested by feeding malicious input to it and then observing its response. This procedure will divulge backdoor vulnerability. The system's performance should be monitored after feeding it new data. If the model's accuracy or precision remarkably decreases then this could be a sign of poisoned samples and should be investigated further [3].

## **2. Evasion**

The attack on the integrity is also called evasion attack, in which malicious samples are modified at testing time to prevaricate detection. These evasion attacks generally occur at the time of inference or testing phase after the ML system has finished training instead of changing training data set as in poisoning attack. In this attack the inputs which are carefully crafted with small perturbation are passed to ML model which forces it for making mistake. For such evasion attacks, the attacker needs some understanding about the inner workings of the ML system by which they can craft the harmful input samples correctly [7] [8].

For instance of an evasion attack, if considered an image classifier which is designed to detect and identify objects. An adversary could take an appropriately classified photo of a dog but append some slight amount of noise to the ML system. Such minor modifications could not be seen with naked eye but it cannot be concealed from ML model. Thus due to such noise, it classifies the object improperly and produce wrong output. In above case it could be castle instead of dog [9].

### **Preventive Measures for Evasion attack:**

For protecting the ML model against the evasion attack, the ML system must give training with adversarial samples to strengthen its capability to recognize them and become more durable. Since decisions in model are based on input, the input sanitization needs to be performed on training data set. To reduce the effect of evasion attacks, different ML models can be employed with diverse training data sets. Proper and persistent monitoring of ML models is required to detect potential evasion attack attempts. The data tampering can be avoided by putting training data at secured storage location which can be accessed by only authenticated person [10].

## **3. Inference:**

If adversaries attempt to reverse-engineer a ML model by providing some particular input data for purpose of rebuilding the ML model's training samples, then it is an inference attack on ML model. If considered that some samples contains the sensitive or secret information about any client, then with this attacking technique the attacker could try to capture such data by providing particular inputs to model [11] [12].

---

There are mainly three types of inference attacks, which include membership inference attack, property inference attack and recovery of training data. In the membership inference attack, the attacker try to identify whether or not the specific data record belongs to target model's training set or not which resulting into information leakage. Whereas in the property inference attack, the adversary tries to guess particular properties that the owner wants to keep hidden or does not wants to share. This information could be any financial data, demographic information (age, gender, family structure, etc.). Also recovery of training data includes the reconstruction of training data set by intruder so that it can automatically disclose any sensitive information [3].

#### **Preventive Measures against the inference attack:**

The method of cryptography is suitable for defending the inference attacker. Since the original text is replaced with alternative (i.e. cipher text), it becomes very difficult to understand the original data for intruder. Another method is to eliminate the sensitive information before reaching at ML model [11]. The augmentation of data could be helpful for protection against inference attack, which can be achieved by adding any non sensitive data to the training data set of a model. This process makes task harder for attacker to get the information about specific records in data set. For protection against such attacks the authentication systems must be strong. Only the authorized users should be permitted to access the training data set. Also the ML system should configure to give random output to make prediction of ML model's result more difficult [8].

#### **4. Extraction:**

In an extraction attack, adversaries try to extract information about the ML model or the data which can be used for training and decision making. The intention behind this type of attack is to understand the architecture of ML model and get the sensitive data used at the time of training phase. There are several types of extraction attacks. Among which the model extraction, training data extraction and hyperparameter extraction are common. The attacker when extracts or changes the whole target ML model then it is called as model extraction attack. But when adversary captures the data which can be used for training of ML model, then it is training data extraction attack. However in the hyperparameter extraction attack the attacker aims to identify the key features such as structure, complexity or learning rate of the targeted ML model [6] [13].

#### **Protective measures against extraction attacks:**

To prevent threat actors from replicating the model, the encryption of ML model parameters before deployment can be used. This will make the process of replication more difficult. Also the unique watermarks can be used to prove ownership of model training data. Due to which the data could not easily steal. Adding some noise to the generated output will help in hiding the sensitive patterns. At the time of usage of this sensitive data this noise could be removed so that it will generate desired output. The access control must be forced to restrict access to the ML system or model and its training data so that external attacker does not use it [14].

#### **Conclusion:**

Machine learning is a powerful tool used for a multiple fields of business activities today. Still, they are vulnerable to attacks that alter their behavior, which causes disruption to systems and reduces their accuracy and performance. Also it can raise questions over reliability of the machine learning system. The main aim of intruder is to affect machine learning models by

variety of attacks to misguide it through malicious input. Thus the ML models must be robust and must be trained for such malicious inputs. Also the cryptography technique is suitable for intrusion over network. The methods of adversarial training, robust optimization, feature selection, game theory will help to implement necessary and appropriate defenses against those attacks. But still there is not a proper mechanism for protecting from those attacks. Thus such technology should be invented [15].

## References:

1. Qifang Bi, Katherine E. Goodman, Joshua Kaminsky, and Justin Lessler, "What is Machine Learning? A Primer for the Epidemiologist", *American Journal of Epidemiology*, Vol. 188, No. 12, DOI: 10.1093/aje/kwz189, August 14, 2019.
2. Yizheng Xu, "Application Research Based on Machine Learning in Network Privacy Security", 978-1-7281-9837-8/20, 2020 IEEE, DOI 10.1109/CIBDA50819.2020.00060|2020.
3. Maria Rigaki, Sebastian Garcia, "A Survey of Privacy Attacks in Machine Learning", *ACM Comput. Surv.* 56, 4, Article 101 (November 2023), 34 pages, 0360-0300/2023/11-ART101 <https://doi.org/10.1145/3624010>, November 2023.
4. Di Cao, Shan Chang, Zhijian Lin, Guohua Liu, Donghong Sun, "Understanding Distributed Poisoning Attack in Federated Learning", 978-1-7281-2583-1/19/, 2019 IEEE, DOI: 10.1109/ICPADS47876.2019.00042| 2019.
5. Jiale Zhang, Junjun Chen, Di Wu, Bing Chen, and Shui Yu, "Poisoning Attack in Federated Learning using Generative Adversarial Nets", 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering, 2324-9013/19/, 2019 IEEE , DOI: 10.1109/TrustCom/BigDataSE.2019.00057|2019
6. Muhammad Maaz Irfan, Sheraz Ali, Irfan Yaqoob, Numan Zafar, "Towards Deep Learning: A Review on adversarial Attack", 2021 International Conference on Artificial Intelligence (ICAI) | 978-1-6654-3293-1/20/ ©2021 IEEE | DOI: 10.1109/ICAI52203.2021.9445247 | 2021.
7. Jinyuan Jia and Neil Zhenqiang Gong, "Defending against Machine Learning based Inference Attacks via Adversarial Examples: Opportunities and Challenges", arXiv: 1909.08526v2 [cs.CR] 19 Sep 2019|2019.
8. Md. Ahsan Ayub, William A. Johnson, Douglas A. Talbert, and Ambareen Siraj, "Model Evasion Attack on Intrusion Detection Systems using Adversarial Machine Learning", 2020 54th Annual Conference on Information Sciences and Systems (CISS), 978-1-7281-4085-8/20/ , 2020 IEEE|2020.
9. Zeinab Khorshidpour, Sattar Hashemi, Ali Hamzeh, "Learning a Secure Classifier against Evasion Attack", 2016 IEEE 16th International Conference on Data Mining Workshops, 2375-9259/16, 2016 IEEE, DOI 10.1109/ICDMW.2016.46| 2016.
10. Md. Ahsan Ayub, William A. Johnson, Douglas A. Talbert, and Ambareen Siraj, "Model Evasion Attack on Intrusion Detection Systems using Adversarial Machine Learning"2020 54th Annual Conference on Information Sciences and Systems (CISS) 978-1-7281-4085-8/20/ ©2020 IEEE 10.1109/CISS48834.2020.1570617116 | 2020.
11. Stacey Truex, Ling Liu, Mehmet Emre Gursay, Lei Yu, and Wenqi Wei, "Demystifying Membership Inference Attacks in Machine Learning as a Service", 1939-1374 , 2019 IEEE, 10.1109/TSC.2019.2897554 |2019.
12. Milad Nasr, Reza Shokri, Amir Houmansadr, "Machine Learning with Membership Privacy using Adversarial Regularization", ACM ISBN 978-1-4503-5693-0/18/10, <https://doi.org/10.1145/3243734.3243855>| 2018.
13. Tatsuya Takemura, Naoto Yanai, Toru Fujiwara, "Model Extraction Attacks on Recurrent Neural Networks", *Journal of Information Processing*, DOI: 10.2197/ipsjip.28.1010, Vol.28 1010–1024 (Dec. 2020)|2020.
14. Hailong Hu, Jun Pang, "Stealing Machine Learning Models: Attacks and Countermeasures for Generative Adversarial Networks", ACM ISBN 978-1-4503-8579-4/21/12., <https://doi.org/10.1145/3485832.3485838>|2021.
15. Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Mirco Marchetti, "Addressing Adversarial Attacks Against Security Systems Based on Machine Learning", 2019 11th International Conference on Cyber Conflict, NATO CCD COE Publications, Tallinn|2019.

## Artificial Intelligence: Advanced Analysis and Design

**Miss.Shrutika Ramesh Dharamkar**  
MSC II year[comp. science]  
Department Of Computer Science  
Vidya Bharti Mahavidyalaya Amravati  
Shrutikadharamkar243@gmail.com

**Miss.Gopika Shyambabu Ghare**  
MSC II year[comp. science]  
Department Of Computer Science  
Vidya Bharti MV Amravati  
gopikaghare23@gmail.com

**Miss.Aparna R. Sapate**  
Dept. Of Computer Science  
Vidya Bharti MV. Amravati  
aparnasapate268@gmail.com

### Abstract:

Artificial Intelligence (A.I.) is a multidisciplinary field whose goal is to automate activities that presently require human intelligence. Recent successes in A.I. include computerized medical diagnosticians and systems that automatically customize hardware to particular user requirements. The major problem areas addressed in A.I. can be summarized as Perception, Manipulation, Reasoning, Communication, and Learning. Perception is concerned with building models of the physical world from sensory input (visual, audio, etc.). Manipulation is concerned with articulating appendages (e.g., mechanical arms, locomotion devices) in order to effect a desired state in the physical world. Reasoning is concerned with higher level cognitive functions such as planning, drawing inferential conclusions from a world model, diagnosing, designing, etc.

### Introduction:

I have chosen this topic to spotlight on one of the most technological trend these days known as AI (Artificial Intelligent). Therefore; I will discuss some of the most important aspects related to AI in which it will help in a better understanding of Artificial Intelligent and both its advantages and disadvantages to be able to protect ourselves from the upcoming technological trend. This paper will also discuss some of the algorithms used in AI systems.

### History of Artificial Intelligence:

Artificial Intelligence was first proposed by John McCarthy in 1956 in his first academic conference on the subject. The idea of machines operating like human beings began to be the center of scientist's mind and whether if it is possible to make machines have the same ability to think and learn by itself was introduced by the mathematician Alan Turing. Alan Turing was able to put his hypotheses and questions into actions by testing whether "machines can think"? After series of testing (later was called as Turing Test) it turns out that it is possible to enable machines to think and learn just like humans. Turing Test uses the pragmatic approach to be able to identify if machines can respond as humans. ("Smith", (nod))

### Description Artificial Intelligence:

Artificial Intelligence is: the field of study that describe the capability of machine learning just like humans and the ability to respond to certain behaviors also known as (A.I.). The need of Artificial Intelligence is increasing every day. Since AI was first introduced to the market, it has been the reason of the quick change in technology and business fields. Computer scientist is predicting that by 2020, "85% of customer interactions will be managed without a human". ("Gartner", (nod)) This means that human's simple request will depend on computers and artificial intelligence just like when we use Sire or Galaxy to ask about the weather temperature. It is very important to be prepared for AI revelation just like UAE have by installing a state minister for AI in Dubai.

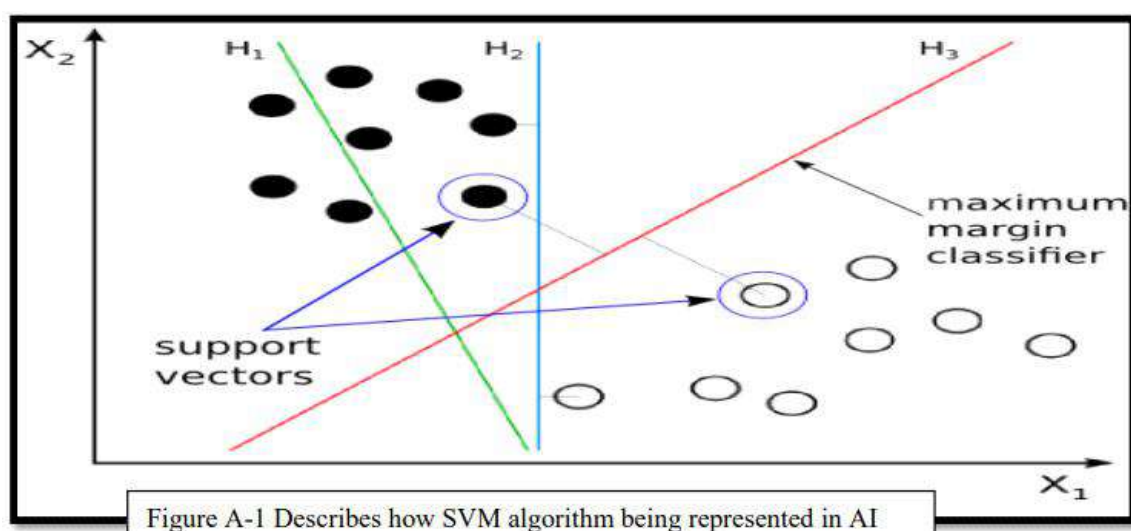
### Pros and Cons of Artificial Intelligence:

AI offers reliability, cost- effectiveness, solve complicated problems, and make decisions; in addition, AI restrict data from getting lost. AI is applied nowadays in most fields whether business or engineering. One of the great tools in AI is called “reinforcement learning” which is based on testing success and failure in real life to increase the reliability of applications. Unfortunately, AI is limited with its capability and functionality. (“Sade”, (nod)) Although Artificial Intelligence made our lives much easier and saved us more time than ever, scientists are predicting that by the huge dependency on AI humanity could extinct. Scientists argue that by having an AI machines, people will be jobless and that will conclude in losing the sense of living. Since machines are learning and doing things more efficiently and effectively in a timely manner, this could be the reason of our extinction.

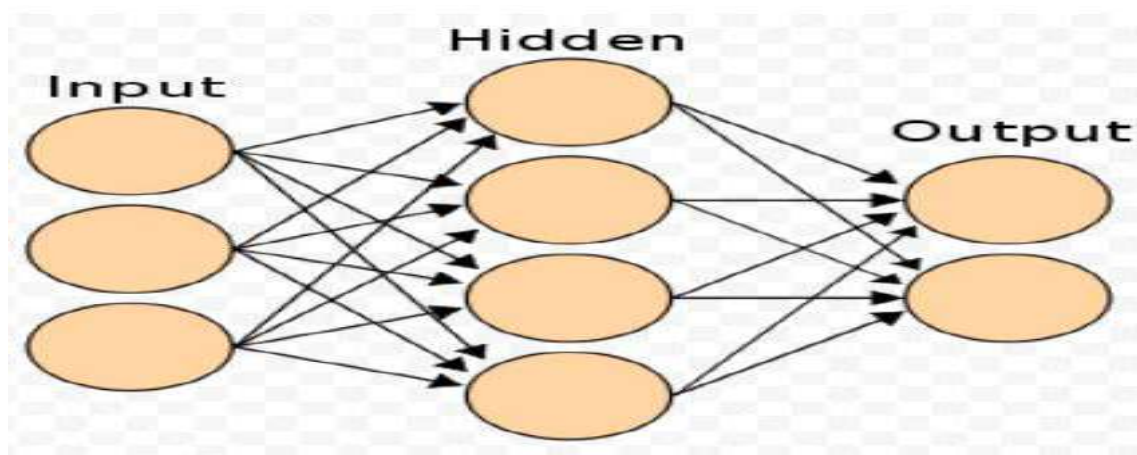
### AI Algorithms and Models:

AI is mainly based on algorithms and models as a technique which is designed based on scientific findings such as math, statistics, and biology (Li& Jiang, (nod)). AI works based on several models such as: Ant Colony Algorithm, Immune Algorithm, Fuzzy Algorithm, Decision Tree, Genetic Algorithm, Particle Swarm Algorithm, Neural Network, Deep Learning and in this report, I will discuss some of the most known models which are: Support Vector Machine, and the Artificial Neural Network.

- Support Vector Machine (SVM) where it is used to build a classification model by Finding an optimal hyper plane based on a set of training examples as shown in (figure A-1) it is also has been used for pattern classification and trend prediction lots of Applications for instance: power transformer fault diagnosis, disease diagnosis and Treatment optimization (Li& Jiang, (nod))



- Artificial Neural Network (ANN) is a representative model of understanding thoughts And behaviours in terms of physical connection between neurons ANN has been used To solve variety of problems through enabling the machine to build mathematical Models to be able to imitate natural activities from brains perspective as shown in (Figure A- 2). By using this algorithm, the machine will be able to identify the Solution of any problem just like human’s brain



### Some Applications on Artificial Intelligence:

AI can be designed using lots of algorithms. These algorithms help the system to determine the expected response which will basically tell the computer what to expect and work accordingly. Here are some of the greatest AI applications that we are probably using in our daily life without knowing:

- Voice recognition
  - Virtual agents:
  - Machine learning platform
  - Ai optimized hardware
  - Decision management
  - Deep learning platform
  - Bio matters
  - Robotic process automation
  - Text analytics and NLP
  - Adaptive Manufacturing :
- Machines that are “able to learn a multitude of tasks from demonstrations, just like their human counterparts can (“You”, 2017))

### AI Design Models:

AI application are a lot around us and in this paper, I will discuss some of the most common application of AI that we always use nowadays which is Virtual Assistants such as Sire, Cortina...etc. Over the past few years smart assistants are becoming a very common technology in most of the smart devices and most importantly, that these assistants are getting smarter than ever. In addition to the awesome help they provide us with, is that every one of these apps has unique features. Artificial Intelligence works according to the following phases: getting the data, clean/manipulate/ prepare the data, train model, test data, and improve the data as mentioned in (figure A-3). Before accessing the data, a business must verify the quality of the data to ensure that it meets the requirement.

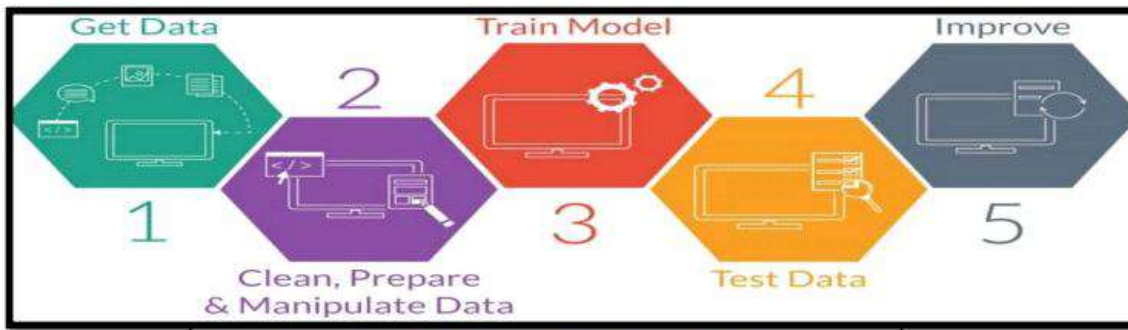


Figure A- 3 Describes Phases of Developing Artificial

**Sire Virtual Assistant:**

Sire is the well-known virtual assistant which uses voice recognitions and typed command in order to perform a certain task within a device. Sire is considered one of AI most used applications. The application simply takes the input from the user such as (e.g. Call dad) and try to find the most related keywords used in this command. Sire tries to eliminate inconsistent result through using the language pattern recognizer and from there to active ontology by searching through the contacts, then it tries to relate the contact named “Dad” and perform the task which is in this case is “Calling” and finally the output of this action will be “calling dad” and to consider all the possible situations refer to (figure A-4).

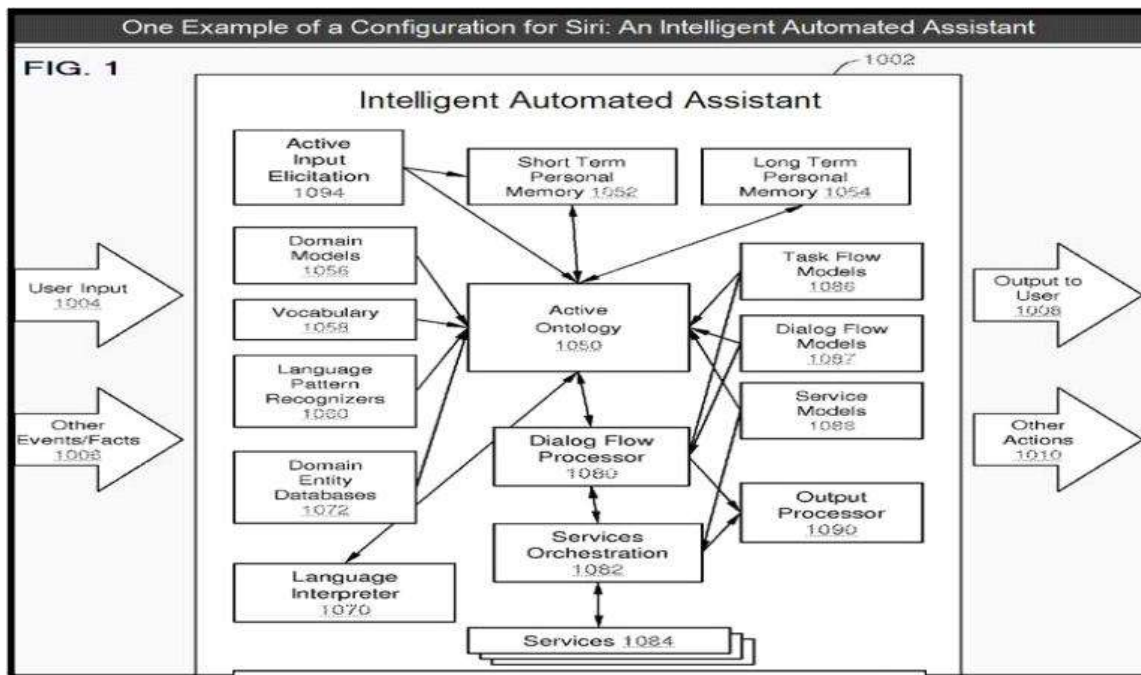


Figure A- 3 Describes one Example of configuration for Siri

In another scenario the architecture of the virtual assistant is shown in (figure A – 5) as we can see the flow of the system starts by taking the input from the user, after that the system decide the conversation strategy module to be used which is a respond from the dialog management module, meanwhile a classification module response to an NLP module. Finally, using the conversation history database is used to analyse the knowledge base construction module

which will response back to the domain knowledge based as explained in detail in (figure A-5)

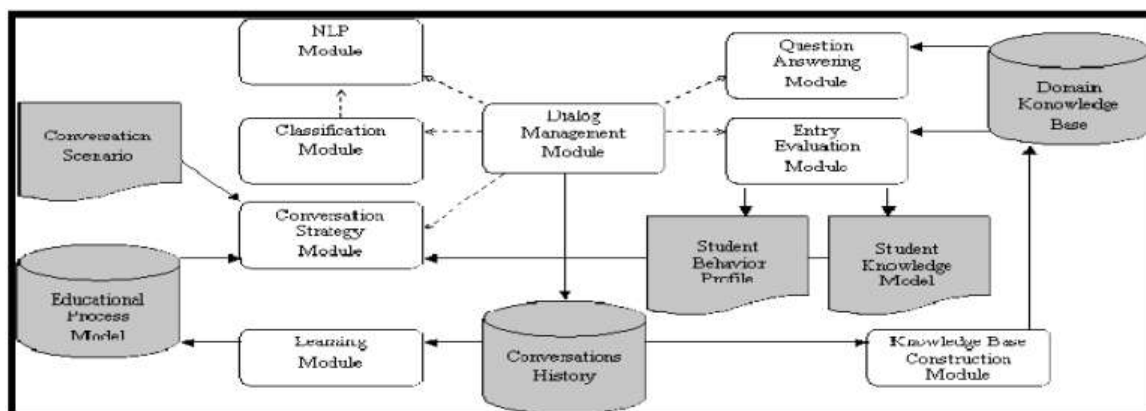


Figure A-5 Describes Proposed conversational agent architecture

### Conclusion:

AI nowadays is being implemented in almost every field of study through several models such as SVM and ANN. We should be able to proceed with knowing and understanding the consequences of every technological trend. In my opinion, we are in the AI revelation era and therefore; we should adopt into this change and welcome it too by embracing AI and moving toward a better society.

### REFERENCES

- 1] Artificial Intelligence Technology and Engineering Applications (2017) ACES JOURNAL, 32, 5th ser., 381-386. Retrieved November 23, 2017, from [file:///C:/Users/lenovo/Desktop/ContentServer%20\(1\).pdf](file:///C:/Users/lenovo/Desktop/ContentServer%20(1).pdf)
- 2] Apple introduces us to Sire, the Killer Patent. (2012, January 19). Retrieved November 25, 2017, From <http://www.patentlyapple.com/patently-apple/2012/01/apple-introduces-us-to-siri-the-killer-patent.html> Acceptability of Embodied Conversational Agent in a health care context (nod) Retrieved November 25, 2017, from [http://www.sanpsy.univbordeauxsegalen.fr/Papers/IVA\\_Additional\\_Material.html](http://www.sanpsy.univbordeauxsegalen.fr/Papers/IVA_Additional_Material.html) [End Times Production] (2017, March 13). Sophia the A.I Robot [Video File] Retrieved from <https://www.youtube.com/watch?v=wimUaNqEJyw>
- 3] Galleon, D., & PhD, C. (2017, October 20). Dubai just appointed a "State Minister for Artificial Intelligence" Retrieved November 22, 2017, from <https://futurism.com/dubai-just-appointed-astate-minister-for-artificial-intelligence>
- 4] Hong, K. (nod). Machine Learning with sickie-learn. Retrieved November 22, 2017, from [http://www.bogotobogo.com/python/scikitlearn/scikit\\_machine\\_learning\\_Support\\_Vector\\_Machines\\_SVM.php](http://www.bogotobogo.com/python/scikitlearn/scikit_machine_learning_Support_Vector_Machines_SVM.php)
- 5] Knight, W. (2017, January 04). What to expect of artificial intelligence in 2017. Retrieved November 23, 2017, from <https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/>
- 6] MA hanta, J. (2017, July 10) Introduction to Neural Networks, Advantages and Applications Retrieved November 23, 2017, from <https://towardsdatascience.com/introduction-to-neuralnetworks-advantages-and-applications-96851bd1a207>
- 7] Mahoney, Z. (2017, April 22). Big Data and Artificial Intelligence for Digital Business Retrieved November 25, 2017, from <http://www.immersiveauthority.com/big-data-artificialintelligence-digital-business/>
- 8] McFarlane N. (2017, October 19). The UAE now has a minister of Artificial Intelligence. Retrieved November 22, 2017, from <http://whatson.ae/dubai/2017/10/uae-now-ministerartificial-intelligence/>
- 9] Sade, A. W., & CHOWDHURY, M. (2012, November) Artificial Intelligence Applications to Critical Transportation Issues Retrieved November 24, 2017, from [https://www.researchgate.net/profile/Said\\_Easa/publication/273576102\\_Design\\_and\\_construction\\_of\\_transportation\\_infrastructure\\_onlinepubstrborgonlinepubscircularsec168pdf/links/55097a910cf26ff55f85932b.pdf#page=14](https://www.researchgate.net/profile/Said_Easa/publication/273576102_Design_and_construction_of_transportation_infrastructure_onlinepubstrborgonlinepubscircularsec168pdf/links/55097a910cf26ff55f85932b.pdf#page=14)
- 10] Smith, C., McGuire, B., Huang, T., & Yang, G. (2006, December) History of Artificial Intelligent [Scholarly project] Retrieved November 20, 2017, from <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
- 11] Yao, M. (2017, August 19). Factories of the Future Need AI to Survive and Compete Retrieved November 23, 2017, from <https://www.forbes.com/sites/mariyayao/2017/08/08/industrial-ai-factories-offuture/#30b20565128e>



## **Disconnected Realities: Unveiling the Challenges and Impact of Network Issues in Mobile Computing**

**Nitin Babaraoji Vasu**

Assistant Lecturer,,  
Institute of Management Studies Maha., Warud. (SGBAU, Amravati)  
Warud, Tq.-Warud, Dist.-Amravati[MAH.] INDIA  
[nitinvasu@gmail.com](mailto:nitinvasu@gmail.com)

### **ABSTRACT:**

Mobile computing has revolutionized the way individuals interact with technology, enabling ubiquitous access to information and services. However, the seamless operation of mobile applications heavily relies on stable and efficient network connectivity. This paper aims to explore the multifaceted challenges and profound impacts of network issues in the realm of mobile computing.

Firstly, the paper delineates the diverse range of network issues encountered in mobile computing environments, including latency, packet loss, bandwidth constraints, and network disruptions. These challenges stem from the inherent characteristics of wireless networks, such as mobility, limited bandwidth, and varying signal strengths.

Secondly, the paper examines the ramifications of network issues on the performance, reliability, and user experience of mobile applications. Degraded network conditions can impede data transmission, degrade application responsiveness, and increase energy consumption, thereby diminishing the overall quality of service.

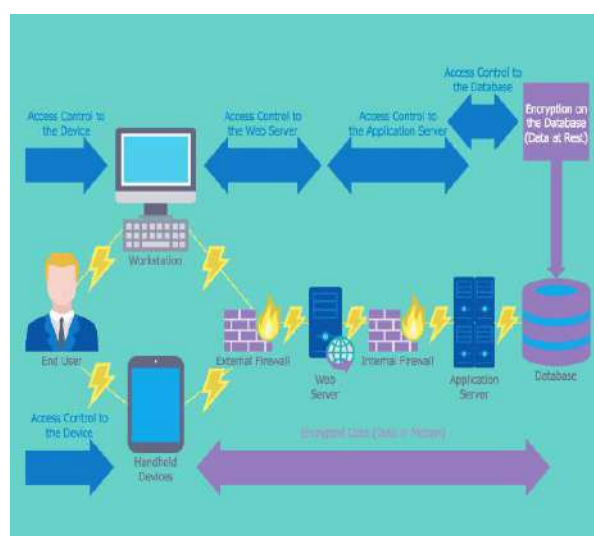
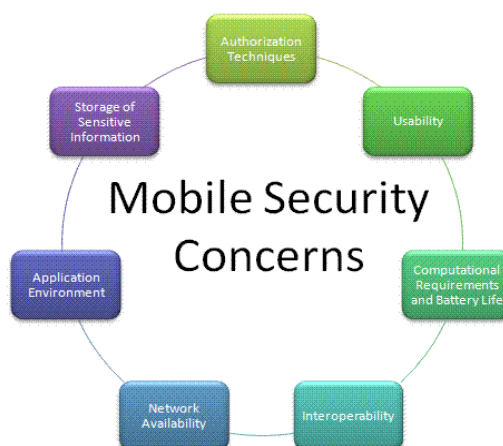
Furthermore, the paper elucidates the strategies and technologies employed to mitigate the impact of network issues in mobile computing. These encompass adaptive bitrate streaming, caching mechanisms, protocol optimizations, and the utilization of offloading techniques to alleviate network congestion and enhance application resilience. Additionally, the paper underscores the significance of proactive network management and fault tolerance mechanisms to preemptively address network anomalies and ensure uninterrupted service delivery in mobile computing environments.

**Keywords:** Network issues, Wireless networks, User experience, Protocol optimizations, Network management, Fault tolerance.

### **1. INTRODUCTION:**

The intersection of network security and mobile computing presents a pivotal arena in today's technological landscape. The proliferation of mobile devices accessing networks introduces numerous security challenges, demanding innovative solutions to safeguard sensitive data and ensure the integrity of communications.

Network security and mobile computing are intricately connected due to the widespread use of mobile devices accessing networks. Ensuring robust network security in the context of mobile computing involves addressing specific challenges associated with the use of smartphones, tablets, and other portable devices that access network resources.



## 2. Key Points:

**Device Security & Mobile Device Management:** Mobile devices need strong security measures such as device encryption, PIN/password protection, biometric authentication, remote wipe capabilities, and secure boot processes to prevent unauthorized access if the device is lost or stolen. Solutions enable centralized management of mobile devices, allowing organizations to enforce security policies, configure settings, and remotely manage applications to maintain security standards across the device fleet.

**Network Access Control** systems help ensure that only authorized and compliant devices can access the network. It checks devices for security compliance before granting access.

**Secure Connectivity:** Using Virtual Private Networks (VPNs) for encrypted connections when accessing corporate networks or sensitive data remotely from mobile devices helps safeguard data in transit.

**Application Security:** Employing mobile application management (MAM) tools to monitor, secure, and manage applications on devices. This includes vetting apps for security before allowing installation and enforcing policies for app behavior.

**Containerization and Segmentation:** Separating personal and work-related data through containerization helps protect sensitive corporate information from potential risks associated with personal usage.

**Mobile Threat Defense (MTD):** Utilizing MTD solutions helps detect and mitigate mobile-specific threats like malware, network attacks, and vulnerabilities targeting mobile devices.

**User Education and Policies:** Training users on best security practices for mobile devices and enforcing strict security policies regarding data access, sharing, and storage on mobile devices is crucial.

**Continuous Monitoring and Updates:** Regularly monitoring mobile devices for security vulnerabilities and ensuring prompt installation of security updates and patches helps maintain their resilience against emerging threats.

As mobile computing continues to evolve, so do the security challenges. Balancing the convenience of mobile access with robust security measures remains a constant challenge for organizations aiming to protect their networks from potential threats stemming from mobile devices.

Network security in the context of mobile computing faces several challenges due to the unique characteristics of mobile devices and their usage.



### 3. CHALLENGES:

Several challenges plague network security in mobile computing. These encompass the susceptibility of wireless networks to interception, unauthorized access due to device loss or theft, malware proliferation through mobile apps, and the complexities of securing diverse devices and operating systems. Moreover, the inherent limitations of mobile hardware and connectivity further complicate security measures.

#### A) SOLVING TECHNIQUES TO ADDRESS CHALLENGES:

To counter these challenges, multifaceted approaches are imperative. Encryption techniques, such as VPNs and secure protocols, fortify data transmissions, rendering them unintelligible to unauthorized entities. Biometric authentication, multifactor authentication, and device encryption bolster device security. Implementation of robust firewalls, intrusion

detection systems, and regular security updates aid in combating malware threats. Additionally, containerization and sandboxing techniques isolate applications, curtailing their ability to compromise the entire system.

#### **B) IMPACT ON DAILY LIFE:**

The ramifications of fortified network security in mobile computing are profound in our daily lives. Individuals experience heightened confidence in conducting sensitive transactions, such as online banking or sharing personal information, fostering a sense of trust in digital interactions. Improved security fosters a more seamless integration of mobile devices in various facets of life, from business operations to healthcare services, augmenting efficiency and accessibility while ensuring the confidentiality of sensitive data.

#### **4. CONCLUSION:**

Network security in mobile computing is an ever-evolving landscape, demanding continuous innovation and adaptation. Addressing these challenges through a combination of advanced technological solutions and user awareness contributes significantly to a safer and more secure digital environment. The resultant impact translates into a more empowered and confident user base, driving further advancements and integration of mobile computing into our daily routines. The exploration of network issues in mobile computing reveals a complex landscape that significantly influences the performance, reliability, and user experience of mobile applications. The challenges posed by wireless networks, including latency, packet loss, bandwidth constraints, and disruptions, underscore the need for proactive strategies and innovative technologies to mitigate their impact.

Throughout this examination, it becomes evident that network issues can hinder data transmission, degrade application responsiveness, and increase energy consumption, thereby diminishing the overall quality of service. However, by employing adaptive bitrate streaming, caching mechanisms, protocol optimizations, and offloading techniques, stakeholders can enhance application resilience and alleviate network congestion.

Moreover, the importance of proactive network management and fault tolerance mechanisms cannot be overstated. By preemptively addressing network anomalies and implementing robust fault tolerance mechanisms, stakeholders can ensure uninterrupted service delivery and enhance the overall user experience in the mobile computing landscape. Moving forward, continued research and development efforts are crucial for advancing the state-of-the-art solutions to address network issues in mobile computing. By fostering collaboration among researchers, industry practitioners, and policymakers, we can collectively navigate the challenges posed by network issues and pave the way for a more resilient and efficient mobile computing ecosystem.

#### **5. REFERENCES**

- [1] [www.google.co.in](http://www.google.co.in)
- [2] <https://www.wikipedia.org/>
- [3] [www.youtube.com](http://www.youtube.com)
- [4] Telecommunication Switching System and Networks, Viswanathan, PHI
- [5] Ad hoc Wireless Networks Architectures, C.Siva Ram Murthy, Pearson

---

## 36

### HTML5 in Web Development

**Varun Sanjay Shende**

Msc 1<sup>st</sup> year 1<sup>st</sup> Sem

Guided By: Dr. B. S. Chinchmalatpure

#### **Abstract :-**

HTML5 is everywhere these days. HTML5 is the new and elegant standard for HTML that provides web users and developers enhanced functionality. The older versions of HTML, HTML 4.01, which came in 1999, and the web development have changed notably since then. HTML 4, XHTML, CSS and the HTML DOM Level 2 are now replaced with HTML5. It was brought to deliver rich content without the need for additional plug-ins and proprietary technologies. The new power of HTML5 supplies the user everything from animation to graphics, music to movies, and can also be used to build complicated web applications and also supports cross-platform. HTML5 standard initiates the development of real-time collaborations in web browsers, which leads to less work for web developers.

Key Words :- Web, users, HTML, HTML5 features, accessibility

#### **INTRODUCTION**

Web is the fastest growing resource that is rapidly and constantly used across almost every platforms. To provide more functionality than the web standards, many software vendors created their own proprietary technique. For e.g. Adobe System Flash, Microsoft silver light, Oracle JAVA FX, Google Gears, Apple's Quick Time etc. web applications run under these proprietary format.

For handling all the jobs that are currently being performed by the proprietary technique W3C is creating a standard on it's latest research on HTML. HTML5 is created by the W3C with Web Hypertext Application Technology Working Group (WHATWG) as a standard that is facilitating the developers and the users without the need of too much additional plug-ins having intensified functionality to increase the platform independence and web openness[1].

#### **HTML5**

As HTML5 is the newer version of HTML, it helps us creating interactive and rich webpages. HTML has grown drastically from simply emphasizing on production of audio, video and animations to providing offline functionality, local storage and geo location on any client side database. The development of HTML5 gives rise to a wide variety of multimedia applications. Without any help of proprietary techniques from the browser it support animations and can play audio and video[2]. For web developer and web designer the new features provided by the HTML5 would add up new values.

HTML5 provides cross platform, which is designed to display webpages on Smart TV, Tablet, PC, Smartphone etc. So many websites as well as browser designers are adopting HTML5 elements. The main temptation for the web developers and browsers is that someone can create rich web pages, web based applications and enhanced forms without mastering or licensing multiple proprietary techniques.

## HTML FEATURES

The new features provided by HTML5 includes

- Audio & Video
- Working Offline
- Drag & Drop
- New Elements
- Canvas-2D/3D Graphics
- Location Based Services
- Web Workers
- New Input Types
- Form Elements

## CANVAS

To combine video and animations on webpages HTML5 uses element for drawing graphics using java script. To present 2D/3D graphics script is used and graphics is contained in the canvas. To make graphics heavy pages render fast, various types of methods for boxes, texts, drawing paths, images.

SVG(Scalar Vector Graphics)

Vector based graphics is defined by the SVG. Even after the compression and enlargement of image the quality of the image will not be lost. The images of SVG can be scalable as well as searched, scripted, indexed and compressed.

## AUDIO & VIDEO

HTML5 provides standards for multimedia files on webpage, where requirement of plug-in for different multimedia files exist. Now nonproprietary formats are fed by the HTML5 in the webpage, which gives a classic way for embedding multimedia files.

The related information is to be controlled as an HTML5 well matched streams and the browser is notified by the `<audio>`,`<source>`and`<video>` tags. These enable users to access the multimedia files without any help of certain players.

```
<audio controls>
```

```
<source src="guitar.mp3"
```

```
type="audio/mpeg">
```

Audio element is not supported by the browser.

```
</audio>
```

```
<video width="480" height="480" controls>
```

```
<source src="afterglow.mp4"
```

```
type="video/mp4">
```

```
</video>
```

---

## LOCATION-BASED SERVICES

User's locations are allotted by the Geo Location Application programming Interface (API). The portable device's geographic location is made obtainable to the web application. By providing assessment with GPS and JS extensions it helps mobile browsers and Location-based applications. The location of user browsing any website can be identified by the HTML5's API, if the user allows to do so.

## WORKING OFFLINE

A brand new procedures for permitting a website or web application which works without network connection are proposed by HTML5. By using cache interface for offline browsing, reducing server load, more speed are great advantages of HTML5. The web applications can be acted as desktop application by Application Cache (AppCache) which allows applications for storing data and programming code.

## WEB STORAGE

Data within user's browser can be stored by a new feature called Web Storage which is provided by HTML5 and it is better than older version of cookies. Client-side SQL database and offline application are supported by this HTML5. Web Storage is more safe and rapid. Without influencing the website's performance high size of data can be stored. Stored data can be accessed through webpage and it stores in pairs of value/name.

## CONCLUSIONS

Latest elements and features are established by HTML5 which permit developers to enhance interoperability, conducting elements in an exact way to save time and costs. HTML5 is an amazing technology, which makes the web even more supreme and substantial, HTML5 has more possibilities i.e. from desktop computers to mobile devices and in domestic appliances. HTML5 has smoothened the line desktop and online applications. Malware Writers that leads to common hacks will be suffered by HTML5 which is available after coming days.

## REFERENCES

1. Research on HTML5 in Web Development 1 Ch Rajesh, 2 K S V Krishna Srikanth 1 Department of IT, ANITS, Visakhapatnam 2 Department of IT, ANITS, Visakhapatnam Ch Rajesh, et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2408-2412
2. HTML5 in Web Development: A New Approach Ashis Kumar Ratha<sup>1</sup>, Shibani Sahu<sup>2</sup>, Priya Meher<sup>3</sup> 1Asst.Prof, Department of Computer Science & Engg., VIT, Bargarh, Odisha, INDIA 2,3 Student Researcher, Department of Computer Science & Engg., VIT, Bargarh Odisha, INDIA
3. HTML5 Web Application Development by Example, J M Gustafson
4. Murach's HTML5 and CSS3 by Zak Ruvalcaba and Anne Boehm
5. Wenling Hu, Hao Yuan, Jiangong Wang, Liang Wang, The Research and Application of Power System Visualization Based on HTML, IEEE 2011. [5] Chen Li-Li, Liu Zheng-Long, Design of Rich Client Web Architecture Based on HTML5, ICCIS, 2012
6. Chen Li-Li, Liu Zheng-Long, Design of Rich Client Web Architecture Based on HTML5, IC

## Itemized Selection of Appropriate Image Steganography Method

**Prof. V. M. Jawade**

PG Department of Computer Science, DCPE, HVPM, Amravati, India. [vjawade03@gmail.com](mailto:vjawade03@gmail.com)

**Prof. R. R. Bhale**

DCPE, HVPM, Amravati, India. [bhalerohan8@gmail.com](mailto:bhalerohan8@gmail.com)

### ABSTRACT

An Image steganography is a research area with lots of algorithms to design an efficient steganography scheme. Implementing the various steganography schemes is an important task to maintain the security in which efficiency of execution needs to be improved. There are a number of algorithms present for the implementation of image steganography, but they do not follow the exact parameters to achieve complete security, so it is necessary to go for the evaluation. Here, the proposed method will select an appropriate steganography scheme according to the imputed image among the existing steganography schemes. This method is proposed to design a robust system which would be itemized as per the requirement of security that needs to be executed.

**Keywords—** Steganography, Stego Ratio, LSB, Encryption, Decryption.

### INTRODUCTION

In steganography the process of hiding information content inside any multimedia content like image, audio, video is referred as an "Embedding" [6]. The process of concealing the secret message in an image file is known as image steganography, hiding the data by taking the cover object as image is referred as image steganography [7]. For increasing the confidentiality of communicating data both the techniques may be combined. There are various steganography techniques used based on the information to be hidden. The various steganography techniques are broadly classified into different categories based on different parameters i.e. accuracy, imperceptibility, embedding capacity, robustness etc. Most used method in this category is least significant bit. But these methods also have some difficulties while implementation. To overcome such problems, improve version of image steganography scheme that is "Itemized Selection of Appropriate Image Steganography Method" is proposed.

### PREVIOUS WORK DONE

In research literature, many LSB methods have been studied to provide various image steganography schemes and improve the performance in terms of Imperceptibility, Embedding Capacity, Robustness. Hedieh Sajedi et al. (2008) [1] has proposed a data hiding scheme that is imperceptible while a big secret image is concealed in a cover image. The main idea is based on dividing the secret image into blocks and considering these blocks as units for embedding. Ramesh Kumar Thakur, et al. (2016) [2] has proposed Various Bits of LSB for Color Images. Deepika Kumar et al. (2015) [3] has proposed method as combination of LSB with the cryptography APIs of Java. Java uses javax.crypto package for encryption and decryption. Dr. Amarendra K, Venkata et al. (2019) [4] has demonstrates striking rationality which is also known as Least Significant Bit (LSB). Tanmay Sinha Roy et al. (2016) [5] has computed Cover Image changes as increase in the LSB Bit substitution, no algorithm is suitable for all types of Steganography.

### PROPOSED METHODOLOGY

The image steganography scheme is an important and difficult task to analyze. Various methods are discussed based on different parameters i.e. accuracy, imperceptibility, embedding capacity, robustness etc. for different image steganography schemes. Image steganography is implemented with various methods like 1-bit, 2-bit, 3-bit (hiding as well as extraction). Choosing the right technique is quite



typical to work out, hence the here proposed "Itemized selection of appropriate steganography method" to select the technique implementation based on imputed image. In the proposed method, the decision-making is done by a system is looking towards the importance of security needed to be executed. The proposed tool also includes the analysis of images using inversion and different color maps, as well as grayscale analysis of images will also work out. The role of the proposed itemized selection of steganography technique comes in existence as the analysis of the image metadata is done, which calculates the stego ratio by using the height, width and actual size of the image. By using the stego ratio, an appropriate steganography method among various LSB methods is chosen for the steganography. Users can share this stego image. On the other hand, the receiver needs a secret key to decrypt the image and get the message hidden in the image.

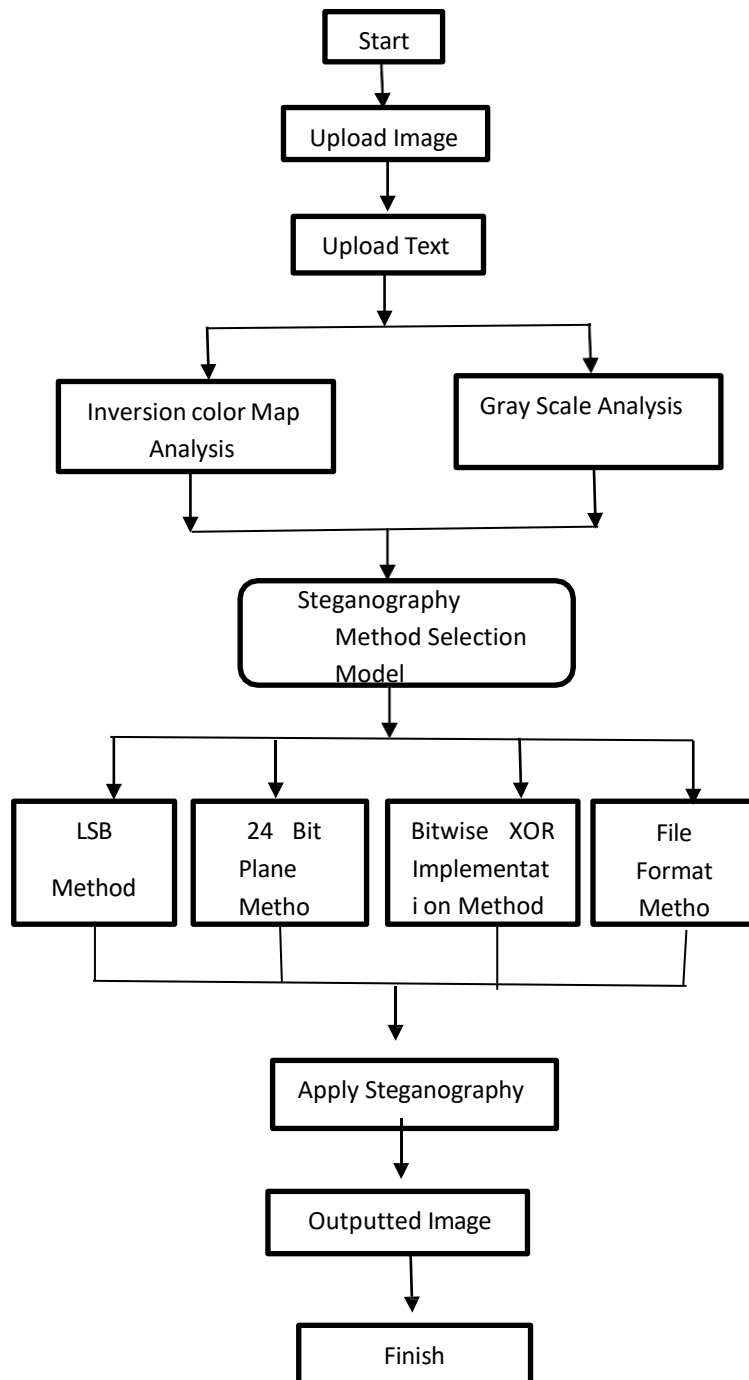


Fig. 1. Flowchart of proposed method

The analysis gives the metadata of image such as actual size of the image, height and width of the image. There are four conditions while selecting the appropriate image steganography method among various LSB bit methods, they are as follows:

- 1) If (actual size > 10 && actual size < 28) then Method = Stego Apply with LSB Method
- 2) If (actual size > 28 && actual size < 56) then Method = Stego Apply with 24 Bit Plane Method
- 3) If (actual size > 56 && actual size < 100) then Method = Stego Apply with Bitwise XOR Implementation Method
- 4) If (actual size > 100) then Method = Stego Apply with File Format Method

#### **OUTCOME AND POSSIBLE RESULT**

This proposed method achieves great results in terms of security by selecting the appropriate steganography method depending upon the imputed image, while, simultaneously it exhibits striking effectiveness in terms of time taken to embed the image with, secret message.

Image File	Width	Height	Actual Size	Time Taken	Steganography Method Used
	194	259	80 kb	19 ms	Bitwise XOR Implementation method
	225	225	112 kb	22 ms	File Format Method
	-1	-1	11 kb	21 ms	LSB Bit Method
	297	169	30 kb	23 ms	24 bit-plane Method
	126	249	242 kb	25ms	File Format Method

**Table 1: Analysis of various stego images.**

## CONCLUSION

This paper discusses various LSB based image steganography methods. But while implementing the various steganography schemes these methods have some problem. It is an important task to choose an efficient algorithm to maintain the security. To overcome such problems new image steganography scheme that is "Itemized Selection of Appropriate Steganography Method" is proposed. In the proposed method the decision making done by system by looking toward the important of security needed to execute. And appropriate steganography scheme is selected according to the imputed image among the existing steganography schemes.

## FUTURE SCOPE

From observations of the proposed method the future work will include no limitation of human visual perception to evaluate the pixel flip ability in larger considered regions. As a further improvement, the model will be implemented advanced steganography techniques. For the encryption process some advanced algorithm may use which enhances the security factor. Also, to compute stego ratio strong mathematical module will be created.

## REFERENCES

- [1] Hedieh Sajedi, Mansour Jamzad, "Cover Selection Steganography Method Based on Similarity of Image Blocks" *IEEE 8<sup>th</sup> International Conference on Computer and Information Technology Workshops*, DOI: 10.1109/CIT.2008.Workshops.34, August, 2008.
- [2] Ramesh Kumar Thakur, Chandran Saravanan, "Analysis of Steganography with various bits of LSB for color images" *IEEE International Conference on Electrical, Electronics, and Optimization Techniques*, 2016.
- [3] Deepika Kumar, Dr. Sanjay Kumar, "Image Based Steganography Using LSB Method and Java Based Encryption" *International Journal of Engineering Trends and Applications*, Vol. 2, Iss. 5, Sep-Oct 2015.
- [4] Dr. Amarendra K, Venkata Naresh Mandhala, B. Chetan Gupta, G. Geetha Sudheshna, V. Venkata Anusha, "Image Steganography using LSB" *International Journal Of Scientific and Technology Research*, Vol. 8, Iss. 12, December 2019.
- [5] Tanmay Sinha Roy, "Image Steganography using LSB Bit-Plane Substitution" *International Research Journal of Engineering and Technology*, Vol. 3, Iss. 12, December 2016.
- [6] Arun Kumar Singh, Juhi Singh, Dr. Harsh Vikram Singh "Steganography in Images Using LSB Technique" *International Journal of Latest Trends in Engineering and Technology*, Vol. 5, No. 1, January 2015.
- [7] Anupriya Arya, Sarita Soni "A Literature Review on Various Recent Steganography Techniques" *International Journal on Future Revolution in Computer Science & Communication Engineering*, Vol. 4, No. 1, January 2018.
- [8] Ressi Dwtias Sari, Andysah Putera Utama Siahaan "Least Significant Bit Comparison between 1-bit and 2-bit Insertion" *International Journal for Innovative Research in Multidisciplinary Field*, Vol. 4, No.10, October 2018.

## **Impact of Skill enhancement Programme on Developing Multimedia Tools for Innovative Learning Solution**

**Ravisha R. Ambekar**

Research Scholar, M.Sc. Extension & Communication, NET  
S.G.B. Amravati University, Amravati  
e-Mail : ravishagadbail@gmail.com, Mob. No. 7219399748

### **ABSTRACT-**

The educational technology can play a major role in revamping the Indian educational system. Multimedia-based education is a combination of interactive and non-interactive learning material with a stimulating, relevant video, animated components and graphics. When education process is creative, problem based, interactive, and target based, learners always interested in learning. Such learning environment could be given by means of Multimedia based Education. An ineffective learning environment could be evidenced that if instructions are designed without considering the multimedia principles. A research is required to identify the elements and structure, which can help the learner to increase the skill level, reduce the learning time and improve the performance of a learner. The study aim to explore the students' perception and digital platforms towards the effective use of multimedia tools through skill enhancement programs for innovative learning solution. The study was conducted to effective use of multimedia tools for enhancing the learning skill. This study has been conducted in secondary school education where instruction and assessments were conducted in multimedia environment. The sample of the study consists of total 120 students of 09th standard. To assess impact of skill development programme on designing and developing multimedia tools for innovative learning solution researcher can be used self-administered skill development test in application of multimedia technology i.e. power point and instructional video . An experimental research design has been selected to conduct the present study. The findings of the study revealed that the multimedia tools was support to develop innovative approach towards learning . Results of the study show that the effectiveness of multimedia embedded classroom was found effective for improved the achievement of the students. Multimedia-based learning is a definite need for the current situation. The paper concluded that Specific skills and Careful Planning, that influence learner perception of multimedia education through skill enhancement programs. The findings may contribute to the existing knowledge of multimedia technology as an educational resource.

Keywords:- Impact, Skill, Developing ,Multimedia Tools, Learning Solution

### **Introduction-**

Multimedia are different courseware designed for teaching and learning and they have been proven effective in accelerating the process of learning. The traditional way of teaching and learning is empowered with the multimedia technology. Multimedia technology also enriches the content of computer based education by providing media rich study materials for students. (Brown, 1995). Learning is a process of all about acquiring new knowledge, sharpening skills, enhances performances, and better understanding. The multimedia based instructional designs provide a platform to learn better, faster and even on self-pace (Weeks C.etal 2006) . It has been observed that learners enjoy studies through computer assisted multimedia instructions, and learn thoroughly (Q. Faryadi Q. 2006). An innovative instructional design is required for better understanding of educational concept (Singh V. K. 2003). Students are able to

concentrate in the class because of the well-designed interactive learning material (Neo T.K etal.2006). Ineffective learning environment could be evidence if instructions are designed without considering the multimedia principles (Moody J. etal 2009). Multimedia Based Learning (MBL) could help in recalling the past learning content and to reconnect with current leaning content( B. Kolloffel etal 2009). To maintain the learning space and interest with students, a learning environment should have adopted various tools and technical techniques (A. Paladino 2008). Use of ICT in education can create new reforms in teaching - learning process in all disciplines of education (Pulkkinen, 2007; Wood, 1995).

### **Literature review-**

According to Andresen & Brink (2013) Multimedia applications can facilitate various Multimedia applications, students can change their focus from acquiring information , analysis and presentation of information. The role of teacher and student have changed the last 25 years. Teachers get new competencies and new roles in a multimedia-learning environment. Multimedia Based Learning resources can create an interactive learning environment than the traditional learning environment (Neo, T. K. etal 2001, Rengarajan.V etal,2007). Multimedia-based learning resources make the learner to enjoy the learning with interactivity (Pang K.2009). Learning process is stumulized when the training programme are given as demonstration method, and these demonstration can be easy by means of digital video and animation ( Junaidu S.2008). Being spent in active learning situation and increase the practices ( Euagelos T.etal 2006). When a class is being conducted for 90 minutes, average involvement of a learner in learning would be 47%. This percentage could be improved by adding interactivity. Multimedia based interactive events could help a learner participate actively in the class and these results in retention( S. G. Smith 1993) . Designing such interaction requires powerful multimedia tools (C. O. Nuallain etal 2006). Using multimedia in the teaching learning environment supports students to become critical thinkers quick learners, and problem-solvers ( Neo, T. K. etal 2001)

Objectives of the Study :-

- 1)To explores the knowledge about digital platforms available for the education system in delivering skill enhancement programs .
- 2) To identify gaps in existing skill and effective use of digital tools of various skill enhancement programs.
- 3) To develop the skill of two identified application of multimedia technology i.e. power point and instructional video

Hypothesis :-

- 1)There will be significant difference between the pre and post test scores of the control and experimental group students on skill development program me of two identified application of multimedia technology i.e. power point and instructional video

Methodology :-

- Research Design :- The pre-test, post-test experimental design was used for this study.
- Sampling and Sampling Techniques :- The population of the study comprises secondary school level students. The sample was selected from schools in Amravati city, State of Maharashtra. Random sampling technique was adopted for selection of the secondary school students. Total 120 students of 09th standard were selected and divided into identical two

groups viz. control and experimental. From each class 60 students were selected and divide into two groups, control group and experimental group.

- **Research Tool Used :-** A self administered skill development test in application of multimedia technology i.e. power point and instructional video for class 09th standard was used for data collection. Skill development test was prepared giving due to consider for their performance has been assessed on 5-point Likert scale where 5 indicated- very satisfied, 4 indicated- satisfied, 3 indicated- neutral, 2 indicated- dis satisfied and 1 indicated- very dissatisfied
- **Statistical Technique Used :-** The data was analyzed with the help of descriptive statistics and 't' test was used.

### **Result and Discussion :-**

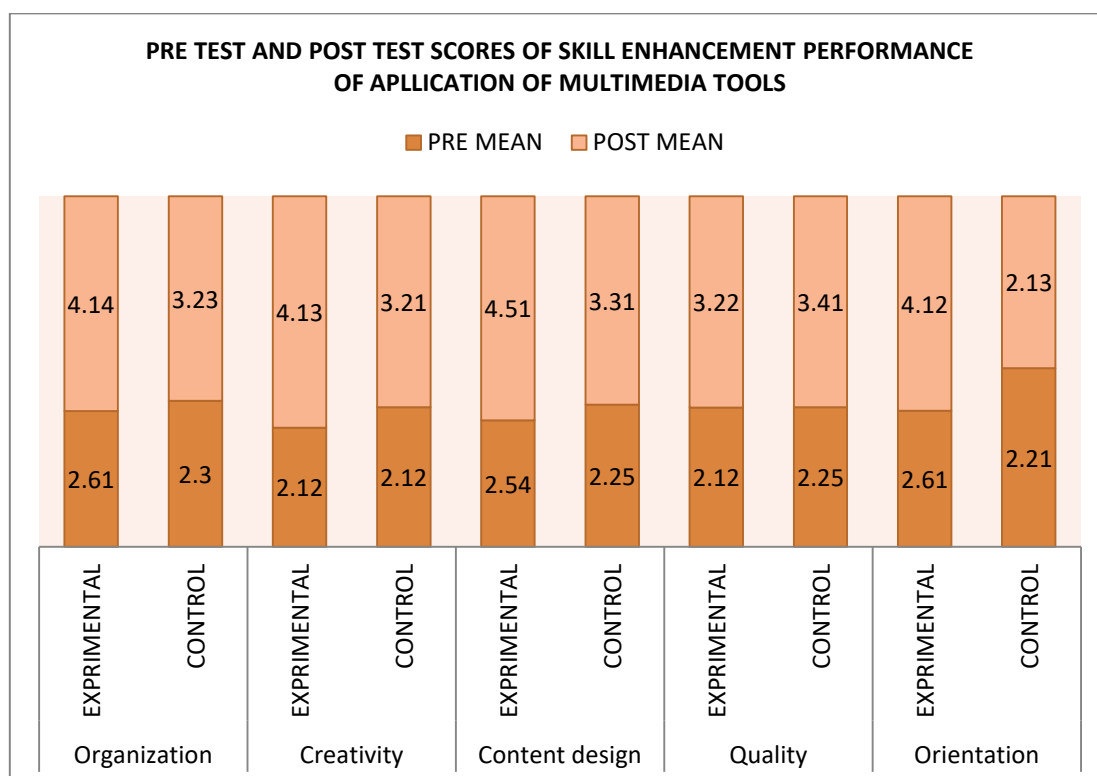
For the present study total 120 secondary school students were selected randomly. Further they were equally divided into Control and experimental groups. The following Table-1 shows that the Comparison of the two groups pre and post test scores related to the skill enhancement performance of application of multimedia technology i.e. power point and instructional video.

**Table- 1 Comparison of the two groups pre and post-test mean scores related to the skill enhancement performance of application of multimedia technology i.e. power point and instructional video.**

ELEMENTS	GROUPS	TEST		t	df		
		PRE MEAN	SD				POST MEAN
Organization	EXPRIMENTAL	2.61	0.51	4.14	0.44	2.41**	59
	CONTROL	2.30	1.00	3.23	0.64	2.40**	59
Creativity	EXPRIMENTAL	2.12	1.23	4.13	0.54	2.16**	59
	CONTROL	2.12	1.14	3.21	1.31	2.40**	59
Content design	EXPRIMENTAL	2.54	0.53	4.51	1.50	2.36**	59
	CONTROL	2.25	1.12	3.31	1.00	2.04**	59
Quality	EXPRIMENTAL	2.12	1.25	3.22	1.43	2.46**	59
	CONTROL	2.25	0.74	3.41	0.91	3.41**	59
Orientation	EXPRIMENTAL	2.61	1.15	4.12	0.84	3.26**	59
	CONTROL	2.21	1.26	2.13	0.46	1.25	59

\*\*p<0.01

**Figure-1 Comparison of the pre and post test mean scores related to the skill enhancement performance of application of multimedia technology i.e. power point and instructional video**



The result of the paired samples t test is presented in the table 1 and figure 1 which shows that both the experimental and control group performed better in their post-test except control group where significant improvement not found. Post test mean scores ranged from 4.12 to 4.51 in the experimental group whereas in the pre test it ranges from 2.12 to 2.61. The P value is less than 0.01 related to five elements of skill development on designing and developing multimedia tools which shows that statistically significant difference in pre test and post-test mean scores between the experimental group related to five elements of skill development on developing multimedia tools.

But Pre test mean scores ranged from 2.12 to 2.25 in the control group whereas in the post test it ranges from 2.13 to 3.41. The P value is less than 0.01 related to five elements (Organization, Creativity, Content design, Quality, orientation) of skill development on developing multimedia tools which shows that statistically significant difference in pre-test and post-test mean scores between the control group related to five elements of skill development on developing multimedia tools. In both Control and experimental groups highest skill improvement has been observed in Organization, Creativity, Content design, Quality, orientation. It clearly indicates that the selected control and experimental groups both are significant difference in the pre-post test score on impact of skill Development programme on designing and developing multimedia tools for innovative learning solution

### Conclusions-

The result of the study indicated that Skill enhancement Programme is beneficial to various subjects at secondary level. Multimedia-based learning environment is strengthening the students learning capabilities. Multimedia tools can improve educational activities. Multimedia technology is form new paradigm in education system

Creating the multimedia-based learning environment useful for solve the learning related issues, understanding related concerns. Multimedia-based environment has the ability to

provide the innovative learning solution for designing the useful content for their respective courses. This is the way by which, we can utilize the optimum benefits of multimedia-based learning in education in near future.

Multimedia-based learning is considered to be an important area and will continue as important learning platform in near future especially in skill based learning programs. Effective implementation of multimedia-based instruction tools could definitely open a new era of learning practices and provide a new paradigm to learners in the days ahead.

The present study concluded that Multimedia-based education has changed teaching-learning process and reforms the traditional method in a innovative pedagogical paradigm.

## References-

1. Andreson, B.B., Brink, K., & UNESCO Institute for information Technologies in Education. 2013 *Multimedia in education: Curriculum*. Moscow: UNESCO Institute for information Technologies in Education.
2. G. Lightbody, P. McCullagh, M. Hutchison, and C. Weeks, "The Supporting Role of Emerging Multimedia Technologies in Higher Education," in 7th Annual Conference, 2006, pp. 46–54.
3. K. Pang, "Video-Driven Multimedia, Web-Based Training in the Corporate Sector: Pedagogical Equivalence and Component Effectiveness," *International Review of Research in Open and Distance Learning*, vol. 10, no. 3, pp. 1–14, 2009.
4. S. Junaidu, "Effectiveness of Multimedia in Learning & Teaching Data Structures Online," *Turkish Online Journal of Distance Education-TOJDE*, vol. 9, no. 4, pp. 97–107, 2008.
5. A. Panagiotis, M. Elias, S. Apostolos, and T. Euangelos, "Multimedia: an instructional tool in the teaching process of al- pine ski .," in *Current Developments in Technology-Assisted Education (2006)*, 2006, pp. 941–945.
6. L. L. Jones and S. G. Smith, "Multimedia technology: A catalyst for change in chemical education," *Pure and Applied Chemistry*, vol. 65, no. 2, pp. 245–249, 1993.
7. Pulkkinen J. (2007). *Cultural globalization and integration of ICT in education* in K. Kumpulainen (Ed.). *Educational technology: Opportunities and Challenges*. Oulu, Finland: University of Oulu, pp. 13-23.
8. Wood D. (1995). *Theory, training and technology Part I*. *Education Train.*, 37(1), 12-16.
9. Q. Faryadi, "Bye, Bye Verbal-only Method of Learning: Welcome Interactive Multimedia," *ERIC-Educational Resources Information Center*, pp. 1–5, 2006.
10. V. K. Singh, "Does Multimedia really improve learning effectiveness?," in *Asia Pacific Conference on Education Re-Envisioning Education: Innovation and Diversity*, 2003, pp. 1–9.
11. B. S. P. Teoh and T.-K. Neo, "Innovative teaching: Using multimedia to engage students in interactive learning in higher education," *IEEE Xplore*, pp. 329–337, 2006.
12. Brown, P.J. (1995), *creating educational hyper documents: can be economic? Innovations in education and training international*, 32(3), August, 201-208
13. G. Krippel, A. J. Mckee, and J. Moody, "Multimedia use in higher education: promises and pitfalls," *Journal of Instructional Pedagogies*, pp. 1–8, 2009.
14. T. H. S. Eysink, T. de Jong, K. Berthold, B. Kolloffel, A. Paladino, "Creating an Interactive and Responsive Teaching Environment to Inspire Learning," *Journal of Marketing Education*, vol. 30, no. 3, pp. 185–188, May 2008.
15. S. Dervan, C. Mccosker, B. Macdaniel, and C. O. Nuallain, "Educational Multimedia," *Current Developments in Technology-Assisted Education*, pp. 801–805, 2006.
16. Damodharan and Rengarajan, V, "Innovative Methods of Teaching," *ebookbrowse*, pp. 1–16, 2007.
18. M. Opfermann, and P. Wouters, "Learner Performance in Multimedia Learning Arrangements: An Analysis Across Instructional Approaches," *American Educational Research Journal*, vol. 46, no. 4, pp. 1107–1149, Aug. 2009.



## IoT-based Automated Smart Irrigation System Using Sensors for Farming

**Dr. Avinash B. Kadam<sup>1</sup>, Mr. Chandrakant R. Patorkar<sup>2</sup>, Miss. Shubhangi D. Falke<sup>3</sup>**

<sup>1</sup>Assistant Professor, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

Corresponding author: E-mail:avinashkadam28@gmail.com<sup>1</sup>

<sup>2</sup>Research Scholar, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

E-mail:chandrakantpatorkar7@gmail.com<sup>2</sup>

<sup>3</sup>Research Scholar, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Maharashtra 443201, India

E-mail:shubhangifalke96@gmail.com<sup>3</sup>

### ABSTRACT

The IoT-based Automated Smart Irrigation System is designed to enhance the efficiency and effectiveness of irrigation practices in agriculture through the integration of cutting-edge technologies such as Internet of Things (IoT) and sensor networks. By incorporating a network of sensors, actuators, data processing units, and communication modules, this system aims to provide farmers with real-time information and control over irrigation processes, resulting in optimal water usage, increased crop yield, and reduced resource wastage.

**Keywords:** IoT, Smart Irrigation, Sensors, IoT Gateway, Edge Computing, Cloud Based.

### I. INTRODUCTION

The IoT-based Automated Smart Irrigation System is designed to enhance the efficiency and effectiveness of irrigation practices in agriculture through the integration of cutting-edge technologies such as Internet of Things (IoT) and sensor networks.

The proposed system uses Arduino IDE, various sensors, Wi-Fi, and GSM modules, an LCD display, and a DC motor to collect and analyze data related to soil moisture, crop growth rate, water level, and animal intervention. The system sends real-time alerts to the farmer via SMS when any abnormal condition is detected [1].

### II. LITERATURE REVIEW

The Literature Review of IoT-based Automated Smart Irrigation System is designed to enhance the efficiency and effectiveness of irrigation system in agriculture. In this automated Irrigation system IoT-based technologies, challenges and outcomes in Agriculture. The Critical Data Analysis of Various IoT-based Automated Smart Irrigation System to develop farming techniques for increasing the Crops Yields for Human. We have various paper surveys on smart irrigation System in farming for sustainable water and energy for Crops Cultivations.

The system comprises various sensors, microcontrollers, wireless communication, and other components that work together to provide real-time data on the environmental conditions of crops. With the use of IoT technology, farmers can reduce their water usage, minimize waste, and increase their crop yield, contributing to global food security and sustainability [1]. In this project, only soil moisture content is controlled as other parameters are hard to control on an agriculture field. Also, for remote monitoring of the sensors, the Thing speak cloud server is used. In this method, the overall cost of equipment is tried to minimize [2]. In this model in the results are analyzed for multiple ML techniques and the results of GBRT based approach are very encouraging. Such kind of techniques could help in achieving optimum utilization of

precious fresh water resources in irrigation, which is a need of the hour in many water stressed countries like India [3]. In practice, we have succeeded in measuring meteorological parameters such as temperature, humidity, percentage of rain and light intensity, thus visualizing the electrical parameters of a photovoltaic installation. With artificial intelligence and also on the Internet of Things, the latter has enabled us to solve major problems in the field of agriculture, such as the waste of water and the quality of products [4]. They propose a smart irrigation model-based IoT system that gradually learns the watering nature of a plant without any pre-prepared data initially given to it. As a proof of concept, we implemented a prototype application. This application adapts itself to the conditions necessary for irrigation after a couple of manual irrigations. To evaluate its performance, we devised tests both for manual and automatic irrigation when different ML algorithms are used [5]. The proposed work can be used in various irrigation models like lateral move irrigation, surface irrigation, sprinkler irrigation and drip irrigation [6]. In this paper we have Critical Data Analysis of Various IoT-Based Technologies for Automated Smart Farming [7]. In this paper they are used IoT-Based Technologies for Automated Irrigation System for Agriculture [8].

### III. METHODS AND MATERIALS

An IoT-based Automated Smart Irrigation System using sensors for farming involves the conceptual representation of the various components, technologies, and processes involved in creating an intelligent irrigation system that leverages Internet of Things (IoT) devices and sensors. The IoT-based Automated Smart Irrigation System using sensors for farming represents a technological leap in agricultural practices by integrating IoT devices and sensors to enable data-driven and efficient irrigation management. This Smart Irrigation system is the fundamental components, workflow, benefits, challenges, and potential of such a system in revolutionizing modern farming techniques.

#### 1. Components Sensors:

- **Soil Moisture Sensors:** Measure soil moisture levels to determine when irrigation is needed.

**Weather Sensors:** Collect data on temperature, humidity, wind speed, and solar radiation to adjust irrigation schedules based on environmental conditions.

- **Rainfall Sensors:** Detect rainfall and adjust irrigation schedules accordingly to prevent overwatering.

#### Actuators:

- **Solenoid Valves:** Control the flow of water to different areas of the farm based on sensor inputs and irrigation schedules.

#### Data Processing Unit:

- **Microcontroller( Arduino UNO):** Collects data from sensors, processes it, and makes decisions based on predefined algorithms or machine learning models.

#### Communication Modules:

- **Wireless Connectivity**



**(Wi-Fi, LoRa):** Enables seamless communication between

sensors, actuators, and the central control unit.

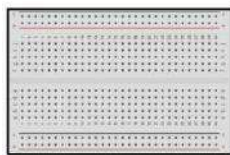
### Other Components:

#### Power Supplies:

- Power sources for sensors, microcontrollers, and communication modules (batteries, solar panels, or a combination)

#### Ultrasonic Sensor:

- An ultrasonic sensor is an instrument that measures the distance to an object using ultrasonic sound waves. An ultrasonic sensor uses a transducer to send and receive ultrasonic pulses that relay back information about an object's.



**Breadboard**



**Jumper Wire**



**Water Pump Motor**



**Display**

## 2. IoT Architecture of SIS Using Sensors for Farming

The development of an IoT-based Automated Smart Irrigation System using sensors for farming involves a structured methodology to ensure successful implementation. Certainly, let's discuss the methodology and materials required for implementing an IoT-based Automated Smart Irrigation System using sensors for farming:

### 1. Sensor Selection and Integration:

- Choose appropriate sensors such as soil moisture sensors, weather sensors, and rain sensors.
- Integrate sensors with microcontrollers or IoT platforms.

### 2. Hardware Setup:

- Assemble the hardware components, including sensors, microcontrollers (Arduino), and communication modules (Wi-Fi, LoRa, etc.).
- Design the physical layout for sensor placement across the farm.

### 3. Data Processing and Decision-Making:

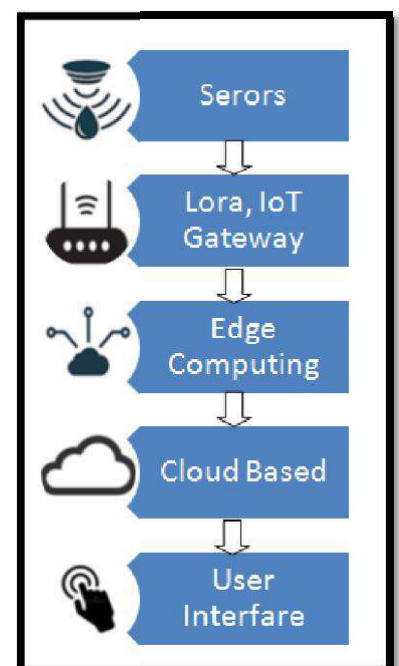
- Develop algorithms to process sensor data and make irrigation decisions.
- Incorporate data from weather forecasts and historical data.

### 4. Central Control Unit Development:

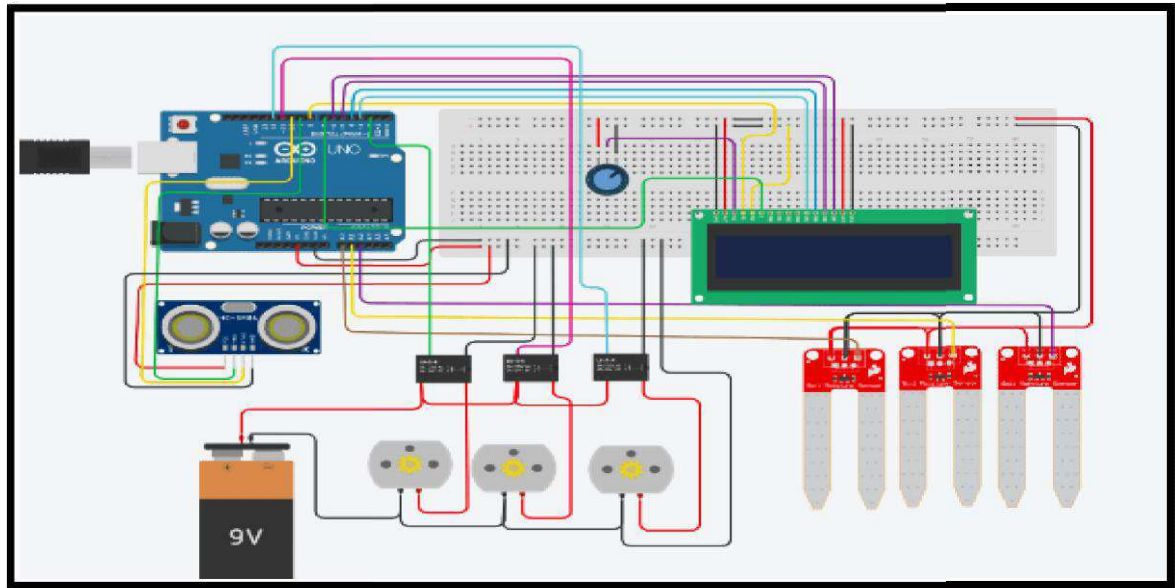
- Build the central control unit, either a cloud-based platform or server.
- Develop software to receive and process sensor data and send commands to actuators.

### 5. User Interface Development:

- Create a user interface such as a web application or mobile app for farmers to monitor and control the system.
- Implement features for setting irrigation schedules and receiving notifications.



**Fig. 2.1 | IoT Architecture of SIS Using Sensors for Farming**



### 3. Benefits:

- **Water Efficiency:** The system ensures that water is applied only when necessary, reducing water wastage and associated costs.
- **Increased Yield:** Optimal irrigation leads to healthier plants and higher crop yields.
- **Resource Savings:** Efficient water use and reduced energy consumption contribute to sustainability.

### 4. Challenges:

- **Sensor Calibration:** Ensuring accurate sensor readings requires regular calibration.
- **Network Reliability:** transmission. Dependable connectivity is crucial for real-time data

## IV. APPLICATIONS

The application of an IoT-based Automated Smart Irrigation System using sensors for farming is aimed at revolutionizing agricultural practices by introducing data-driven, efficient, and sustainable irrigation management. Here are some key applications and benefits of such a system:

### 1. Precision Irrigation

Sensors continuously monitor soil moisture levels, weather conditions, and other relevant parameters. This data is used to tailor irrigation schedules and amounts to the specific needs of different crops and soil types.

### 2. Water Conservation

By irrigating only when necessary and optimizing water usage, the system reduces water wastage, helps conserve this practices.

### 3. Increased Crop Yield

Valuable resource, and contributes to sustainable farming

Providing crops with the right amount of water at the right time enhances plant health and growth, leading to higher yields and improved crop quality.

### 4. Energy Efficiency

Automated systems ensure that irrigation occurs during optimal times, reducing energy consumption associated with pumping water. This is especially important in regions with limited energy availability.

## V. CONCLUSION & FUTURE SCOPE

The IoT-based Automated Smart Irrigation System using sensors for farming represents a technological leap in agricultural practices by integrating IoT devices and sensors to enable efficient irrigation management system. This system offers numerous benefits, including precision irrigation, water conservation, increased crop yields, energy efficiency, and reduced labour demands. By leveraging real-time data from soil moisture sensors, weather sensors, and other sources, farmers can make informed decisions about when and how much to irrigate, optimizing resource usage and minimizing wastage.

Furthermore, the potential for remote monitoring and control through user-friendly interfaces empowers farmers to manage their irrigation systems conveniently, regardless of their location. The data-driven insights gained from the collected information provide a deeper understanding of crop behaviour, enabling continuous improvement of irrigation strategies over time. As technology continues to evolve, the future of IoT-based smart irrigation systems holds the promise of even greater advancements, such as the integration of artificial intelligence, autonomous operation, and enhanced collaboration among farmers.

In a world facing the challenges of climate change, population growth, and resource constraints, the application of smart irrigation systems represents a significant step forward in sustainable agriculture. By embracing these innovations, farmers can contribute to increased productivity, efficient resource utilization, and a more resilient food supply, ultimately shaping a more sustainable and prosperous future for agriculture and the planet as a whole.

## VI. REFERENCES

- [1] A. K. D B, D. N, B. Bairwa, A. K. C S, G. Raju and Madhu, "IoT-based Water Harvesting, Moisture Monitoring, and Crop Monitoring System for Precision Agriculture," 2023 International Conference on Distribal Computing and Electrical Circuits and Electronics (ICDCECE), Ballar, India, 2023, pp. 1-6, doi: 10.1109/ICDCECE57866.2023.10150893.
- [2] A. Srivastava, D. K. Das and R. Kumar, "Monitoring of Soil Parameters and Controlling of Soil Moisture through IoT based Smart Agriculture," 2020 IEEE Students Conference on Engineering & Systems (SCES), Prayagraj, India, 2020, pp. 1-6, doi: 10.1109/SCES50439.2020.9236764.
- [3] G. Singh, D. Sharma, A. Goap, S. Sehgal, A. K. Shukla and S. Kumar, "Machine Learning based soil moisture prediction for Internet of Things based Smart Irrigation System," 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 2019, pp. 175-180, doi: 10.1109/ISPCC48220.2019.8988313.
- [4] H. Youness, G. Ahmed and B. E. Haddadi, "Machine Learning-based Smart Irrigation Monitoring System for Agriculture Applications Using Free and Low-Cost IoT Platform," 2022 International Conference on Microelectronics (ICM), Casablanca, Morocco, 2022, pp. 189-192, doi: 10.1109/ICM56065.2022.10005419.
- [5] K. Cagri Serdaroglu, C. Onel and S. Baydere, "IoT Based Smart Plant Irrigation System with Enhanced learning," 2020 IEEE Computing, Communications and IoT Applications (ComComAp), Beijing, China, 2020, pp. 1-6, doi: 10.1109/ComComAp51192.2020.9398892.
- [6] M. B. Tephila, R. A. Sri, R. Abinaya, J. A. Lakshmi and V. Divya, "Automated Smart Irrigation System using IoT with Sensor Parameter," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 543-549, doi: 10.1109/ICEARS53579.2022.9751993.
- [7] C. R. Patorkar and A. B. Kadam, "Critical Data Analysis of Various IoT-Based Technologies for Automated Smart Farming", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 2023 Vol. 9, pp.411-437.
- [8] M. Bogdanoff and S. Tayeb, "An ISM-Band Automated Irrigation System for Agriculture IoT," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 2020, pp. 1-6, doi: 10.1109/IEMTRONICS51293.2020.9216351.

## Data Mining Algorithms Analysis on Weka for Disease Classification

Sushilkumar R. Kalmegh and Prerna S. Tayade

PG Department of Computer Science, Sant Gadge Baba Amravati University,  
Amravati (M.S.) 444 602, India

### ABSTRACT

In this paper data mining approaches for disease classification are presented. The Weka for disease classification using data mining algorithms is described in this paper. The data preprocessing and algorithm evaluation using Weka is described in this paper. The data mining algorithms ZeroR, Linear Regression, Multilayer Perceptron, Random Forest, Simple K-Means, Hierarchical Clustering, and Farthest First are proposed for disease classification using Weka. The algorithms are evaluated using Weka. These algorithms can be used to classify any disease. These algorithms are compared on the basis of accuracy and disease classification time.

*Keywords: Algorithm Analysis, Comparative Analysis, Data Mining, Disease Classification, Weka*

### INTRODUCTION

As the Internet of medical Things emerge in the field of medicine, the volume of medical data is expanding rapidly and along with its variety. There are several data mining algorithms commonly used for disease classification. Decision trees are used to partition data into subsets based on different features, ultimately leading to a classification. Random forest is ensemble learning method combines multiple decision trees to improve accuracy and reduce overfitting. Support Vector Machines (SVM) is effective in separating classes by finding the hyperplane that maximally separates the data points. K-Nearest Neighbors (KNN) classifies data points based on the majority class among their k-nearest neighbors in the feature space. Naive Bayes algorithm is based on Bayes' theorem and assumes independence between features. It's particularly useful for text classification. Neural Networks, Deep learning techniques, like artificial neural networks, can be applied for disease classification tasks, especially with the availability of large datasets. Despite its name, logistic regression is a classification algorithm suitable for binary and multiclass problems. Besides Random Forest, other ensemble methods like AdaBoost or Gradient Boosting can be effective in combining the strengths of multiple models. The choice of algorithm depends on factors such as the nature of the data, the size of the dataset, and the specific characteristics of the disease being studied. Keep in mind that these algorithms may be used individually or in combination, depending on the complexity of the classification task.

Weka is a popular and versatile open-source data mining and machine learning software. It provides a graphical user interface for a wide range of data preprocessing, clustering, classification, regression, and visualization tasks. For data preprocessing load dataset into Weka. Explore and preprocess data using Weka's tools. This might involve handling missing values, converting categorical variables, or normalizing numeric attributes. Choose a disease classification algorithm. Weka supports a variety of algorithms, including decision trees, support vector machines, k-nearest neighbors, and more. Split dataset into training and testing sets. Use the Explorer interface to select a classification algorithm and configure its parameters. Apply the algorithm to the training set to train the model. Apply the trained model to the testing set. Use Weka's evaluation tools to assess the performance of the model. This may include metrics like accuracy, precision, recall, and F-measure. Experiment with different algorithms

and parameter settings to optimize the performance of model. Weka supports various cross-validation techniques to assess the generalization performance of model. Weka provides visualization tools to help you understand and interpret the results of classification experiments. Once satisfied with the model, save it for future use and apply it to new data when needed. Remember that the specific steps may vary based on the disease dataset and the classification algorithm choose. Weka's user-friendly interface and extensive documentation make it a great tool for both beginners and experienced practitioners in the field of data mining and machine learning for disease classification.

Disease classification involves using machine learning algorithms to analyze and categorize data related to various diseases. The goal is to develop models that can accurately predict or classify the presence or absence of a particular disease based on input features. Disease classification helps organize information in a way that makes sense. By categorizing diseases, we create a systematic way to understand and communicate about them. It provides a common language for healthcare professionals to discuss and share information about diseases. Imagine if everyone had their own way of describing illnesses—it would be chaos! Different diseases often require different approaches to treatment. Classifying them helps healthcare providers determine the most effective interventions based on the shared characteristics of a particular group of diseases.

Scientists and researchers use disease classification to study specific groups of diseases, identify patterns, and develop new treatments. It's like putting puzzle pieces together to see the bigger picture of a disease category. Epidemiology is understanding the distribution and determinants of diseases in populations is essential for public health. Classification helps track the prevalence and incidence of diseases, aiding in the development of preventive measures. Public health planning enables health organizations to allocate resources more efficiently. By knowing which diseases are more prevalent or pose greater risks, public health initiatives can be targeted where they are needed most. Medical students and healthcare professionals learn about diseases in a structured way, starting with the basics and building up to more complex conditions. This systematic approach helps in understanding the interconnectedness of different diseases. In essence, disease classification is the roadmap that guides healthcare professionals, researchers, and policymakers in navigating the vast and complex landscape of human health. It's the key to unlocking understanding, treatment, and prevention.

## LITERATURE REVIEW

A comparative research for the particle swarm optimization (PSO), firefly, cuckoo, and bat algorithms based on both synthetic and real medical data sets is presented by [Gong et al., 2019]. Utilizing machine learning and data mining [Wee et al., 2022] proposed to automate the process of referrals. An ensemble of machine learning algorithms to perform clinical text mining against the unstructured referral text in order to derive the relationship among the discovered medical terms was proposed and implemented in [Wee et al., 2022]. The neural network-based text classification model that can classify referrals with high accuracy was developed, tested and reported [Wee et al., 2022]. Association Rules Mining (ARM) is a data mining procedure that discovers the relationship among the items in a dataset that are based on rules that can have some measure of interest. It is commonly known as Market Basket Analysis. According to [Ma et al., 2013], the rule to measure the interest between two items, X and Y, from a dataset, N, can be expressed as Support, Confidence and Lift. For the rule of  $X \rightarrow Y$ , the Support indicates the frequency of the itemset in the dataset, the Confidence is the frequency of the rule that is found to be accurate, and the Lift of the rule represents the ratio of the rule's support to the support of X and Y when they are independent.

$$\text{Support}(X \rightarrow Y) = \frac{\text{Frequency}(X, Y)}{N}$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Frequency}(X, Y)}{\text{Frequency}(X)}$$

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X) \cdot \text{Support}(Y)}$$

The Apriori algorithm was proposed by [Ma et al., 2013] and it has a bottom-up approach in identifying the items that are the subset within the dataset that meet the given threshold of transactions [Ma et al., 2013]. A various of machine learning/deep learning algorithms have been applied to medical and healthcare domain including cancer research [Abdar et al., 2020], prediction of sepsis [Kok et al., 2020], chronic pain detection [Bargshady et al., 2019; Bargshady et al., 2020], coronary artery disease research [Nasarian et al., 2020; Zomorodi-Moghadam et al., 2021; Abdar et al., 2019]. A neural network is a group of machine learning algorithms that mimic the organic human brain to recognize a set of data's underlying relationship through the process of minimizing a loss function via experiencing with the learning of weights on the neurons [Albawi et al., 2017; Abiodun et al., 2018].

In order to deal with issues such as poor large-scale data mining in the medical field, [Jin et al., 2017] proposed a new data collection and mining strategy, analyzed the relevance of obstetric information data in current medical institutions based on Apriori algorithm, focused on the association between cesarean section and the existing signs and the drugs used, and analyzed the association between maternal hospitalization and the time and number of births. A novel multi-source medical data integration and mining solution is presented by [Zhang et al., 2020] for better healthcare services, named PDFM (Privacy-free Data Fusion and Mining).

## PROPOSED DATA MINING ALGORITHMS FOR DISEASE CLASSIFICATION

The various data mining algorithms are used in Weka for disease classification. The proposed algorithms are ZeroR, Linear Regression, Multilayer Perceptron, Random Forest, Simple K-Means, Hierarchical Clustering, and Farthest First. These algorithms are explained in following section.

ZeroR is the simplest classification method which relies on the target and ignores all predictors. ZeroR classifier simply predicts the majority category (class). Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification methods. It constructs a frequency table for the target and selects its most frequent value. There is nothing to be said about the predictor's contribution to the model because ZeroR does not use any of them. ZeroR only predicts the majority class correctly. As mentioned before, ZeroR is only useful for determining a baseline performance for other classification methods.

Making predictions with linear regression, given the representation is a linear equation, making predictions is as simple as solving the equation for a specific set of inputs. Let's make this concrete with an example. Imagine predicting weight (y) from height (x). Linear regression model representation for this problem would be:

$$y = B_0 + B_1 * x_1$$

or

$$\text{weight} = B_0 + B_1 * \text{height}$$

Where  $B_0$  is the bias coefficient and  $B_1$  is the coefficient for the height column. Use a learning technique to find a good set of coefficient values. Once found, plug in different height values to predict the weight. For example, let's use  $B_0 = 0.1$  and  $B_1 = 0.5$ . Let's plug them in and calculate the weight (in kilograms) for a person with the height of 182 centimeters.

$$\text{weight} = 0.1 + 0.5 * 182$$

$$\text{weight} = 91.1$$

This was proved almost a decade later by Minsky and Papert, in 1969 and highlights the fact that Perceptron, with only one neuron, can't be applied to non-linear data. The Multilayer Perceptron, was developed to tackle this limitation. It is a neural network where the mapping



between inputs and output is non-linear. A Multilayer Perceptron has input and output layers, and one or more hidden layers with many neurons stacked together. While in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.

In each iteration, after the weighted sums are forwarded through all layers, the gradient of the Mean Squared Error is computed across all input and output pairs. Then, to propagate it back, the weights of the first hidden layer are updated with the value of the gradient. That's how the weights are propagated back to the starting point of the neural network. One iteration of Gradient Descent is mathematically represented as:

$$\Delta_w(t) = -\varepsilon \frac{dE}{d\omega(t)} + \alpha \Delta_w(t-1)$$

$\Delta_w(t)$ : Gradient Current Iteration

$\varepsilon$ : Bias

$E$ : Error

$\omega(t)$ : Weight Vector

$\alpha$ : Learning Rate

$\Delta_w(t-1)$ : Gradient Previous Iteration

This process keeps going until gradient for each input-output pair has converged, meaning the newly computed gradient hasn't changed more than a specified convergence threshold, compared to the previous iteration.

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

- **Step 1:** In Random forest n number of random records are taken from the data set having k number of records.
- **Step 2:** Individual decision trees are constructed for each sample.
- **Step 3:** Each decision tree will generate an output.
- **Step 4:** Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

The approach k-means follows to solve the problem is called Expectation-Maximization. The E-step is assigning the data points to the closest cluster. The M-step is computing the centroid of each cluster. The objective function is:

$$J = \sum_{i=1}^m \sum_{k=1}^K \omega_{ik} \|x^i - \mu_k\|^2$$

where  $\omega_{ik} = 1$  for data point  $x^i$  if it belongs to cluster k; otherwise,  $\omega_{ik} = 0$ . Also,  $\mu_k$  is the centroid of  $x^i$ 's cluster.

It's a minimization problem of two parts. First minimize J with respect to  $\omega_{ik}$  and treat  $\mu_k$  fixed. Then minimize J with respect to  $\mu_k$  and treat  $\omega_{ik}$  fixed. Technically speaking, differentiate J with respect to  $\omega_{ik}$  first and update cluster assignments (E-step). Then differentiate J with respect to  $\mu_k$  and re-compute the centroids after the cluster assignments from previous step (M-step). Therefore, E-step is:

$$\frac{\partial J}{\partial \omega_{ik}} = \sum_{i=1}^m \sum_{k=1}^K \omega_{ik} \|x^i - \mu_k\|^2$$

$$\Rightarrow \omega_{ik} = \begin{cases} 1 & \text{if } k = \arg \min_j \|x^i - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases}$$

In other words, assign the data point  $x^i$  to the closest cluster judged by its sum of squared distance from cluster's centroid. M-step is:

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^m \omega_{ik} (x^i - \mu_k) = 0$$

$$\Rightarrow \omega_k = \frac{\sum_{i=1}^m \omega_k x^i}{\sum_{i=1}^m \omega_k}$$

Which translates to recomputing the centroid of each cluster to reflect the new assignments. Since clustering algorithms including k-means use distance-based measurements to determine the similarity between data points, it's recommended to standardize the data to have a mean of zero and a standard deviation of one since almost always the features in any dataset would have different units of measurements such as age versus income. Given k-means iterative nature and the random initialization of centroids at the start of the algorithm, different initializations may lead to different clusters since k-means algorithm may stuck in a local optimum and may not converge to global optimum. Therefore, it's recommended to run the algorithm using different initializations of centroids and pick the results of the run that that yielded the lower sum of squared distance. Assignment of examples isn't changing is the same thing as no change in within-cluster variation:

$$\frac{1}{m_k} \sum_{i=1}^{m_k} \|x^i - \mu_c^k\|^2$$

Agglomerative Clustering is widely used in the industry and therefore that is the focus in this research work. Divisive hierarchical clustering is simple once the agglomerative type is handled. The algorithm for perform hierarchical classification /clustering is specified below.

- **Step 1:** First, assign all the points to an individual class/ cluster
- **Step 2:** Next, look at the smallest distance in the proximity matrix and merge the points with the smallest distance. Then update the proximity matrix
- **Step 3:** Repeat step 2 until only a single class/ cluster is left.

So, first look at the minimum distance in the proximity matrix and then merge the closest pair of classes/ clusters. Get the merged classes/ clusters after repeating these steps.

Farthest First is a Variant of K means that places each cluster centre in turn at the point furthest from the existing cluster centres. This point must lie within the data area. This greatly speed up the classification in most cases since less reassignment and adjustment is needed.

## EVALUATIONS OF THE PROPOSED DATA MINING ALGORITHMS FOR DISEASE CLASSIFICATION

The proposed data mining algorithms namely ZeroR, Linear Regression, Multilayer Perceptron, Random Forest, Simple K-Means, Hierarchical Clustering, and Farthest First used for disease classification are described in the previous section. This section provides the evaluation of these algorithms. The evaluation is carried out using Weka 3.8.6 on Windows 10 with 4 GB RAM, 1TB HDD and Intel i3 8<sup>th</sup> generation processor. There are two versions of Weka: Weka 3.8 is the latest stable version and Weka 3.9 is the development version. New releases of these two versions are normally made once or twice a year. For the bleeding edge, it is also possible to download nightly snapshots of these two versions. The stable version receives only bug fixes and feature upgrades that do not break compatibility with its earlier releases, while the development version may receive new features that break compatibility with its earlier releases. Weka 3.8 and 3.9 feature a package management system that makes it easy for the Weka community to add new functionality to Weka. The package management system requires an internet connection in order to download and install packages. The dataset used in this research work is in the Comma Separated Value (CSV) format. The data used to train the data mining algorithms and to identify the disease classification results is shown in figure 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	
2	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	
3	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2	
4	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	
5	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	
6	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	
7	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0	
8	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3	
9	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0	
10	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2	
11	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1	
12	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0	
13	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0	
14	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2	
15	44	1	2	120	263	0	0	173	0	0	1	0	7	0	
16	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0	
17	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0	
18	48	1	2	110	229	0	0	168	0	1	3	0	7	1	
19	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0	
20	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0	
21	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0	
22	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0	
23	58	0	1	150	283	1	2	162	0	1	1	0	3	0	
24	58	1	2	120	284	0	2	160	0	1.8	2	0	3	1	
25	58	1	3	132	224	0	2	173	0	3.2	1	2	7	3	

**Figure 1: Sample Dataset used to Train the Data Mining Algorithms**

The dataset shown in figure 1 comprises 300 samples and 14 attributes. In the pre-processing phase the null samples are removed and missing sample values are adjusted as per the relevant sample values present in the dataset instance. It can be observed from the dataset values distribution there are 4 target classes for the disease classification results. These classes are none, mild, moderate, and severe. The pre-processed data is used to train the data mining algorithms for disease classification.

The comparison of ZeroR, Linear Regression, Multilayer Perceptron, Random Forest, Simple K-Means, Hierarchical Clustering, and Farthest First on the basis of accuracy and classification time is presented in table 1. It can be observed that Farthest First is the most computationally efficient algorithm compared to remaining evaluated algorithms. The Simple K-Means has the highest accuracy among the remaining evaluated algorithms. The Hierarchical Clustering has the worst accuracy. While Multilayer Perceptron has more computational complexity.

**Table 1: Comparative Analysis of Data Mining Algorithms for Disease Classification**

Algorithm	Accuracy	Classification Time (Seconds)
ZeroR	93.72%	0.23
Linear Regression	78%	0.94
Multilayer Perceptron	89.24%	3.06
Random Forest	86.98%	1.96
Simple K-Means	94.03%	0.08
Hierarchical Clustering	48.52%	0.62
Farthest First	91.82%	0.02

## CONCLUSION

In this paper the data mining algorithms ZeroR, Linear Regression, Multilayer Perceptron, Random Forest, Simple K-Means, Hierarchical Clustering, and Farthest First are proposed for disease classification using Weka. The algorithms are evaluated using Weka. These algorithms can be used to classify any disease. These algorithms are compared on the basis of accuracy

and disease classification time. The data mining algorithms ranging from highest to lowest accuracy Simple K-Means, ZeroR, Farthest First, Multilayer Perceptron, Random Forest, Linear Regression, and Hierarchical Clustering, has the accuracy of 94.03%, 93.72%, 91.82%, 89.24%, 86.98%, 78%, and 48.52% respectively. If arranged on the basis of the computational complexity lowest to highest Farthest First, Simple K-Means, ZeroR, Hierarchical Clustering, Linear Regression, Random Forest, and Multilayer Perceptron has disease classification time 0.02, 0.08, 0.23, 0.62, 0.94, 1.96, and 3.06. These algorithms can be further evaluated for more diseases. Despite the improved performance for data mining classification modeling, there are several limitations can be addressed in future. The related algorithms can be improved by incorporating the evaluation feedback results for disease classification.

## REFERENCES

- [1] [Gong et al., 2019] Xueyuan Gong, Liansheng Liu, Simon Fong, Qiwen Xu, Tingxi Wen, and Zhihua Liu, "Comparative Research of Swarm Intelligence Clustering Algorithms for Analyzing Medical Data," IEEE Access Special Section on Trends, Perspectives and Prospects of Machine Learning Applied to Biomedical Systems in Internet of Medical Things, Volume 7, pp. 137560- 137569, Oct 2019.
- [2] [Wee et al., 2022] Chee Keong Wee, Xujuan Zhou, Raj Gururajan, Xiaohui Tao, Jennifer Chen, Rashmi Gururajan, Nathan Wee, and Prabal Datta Barua, "Automated Triageing Medical Referral for Otorhinolaryngology Using Data Mining and Machine Learning Techniques," IEEE Access, VOLUME 10, pp. 44531-44548, May 2022.
- [3] [Ma et al., 2013] H. Ma, Y. Hu, and H. Shi, "Fault detection and identification based on the neighborhood standardized local outlier factor method," Ind. Eng. Chem. Res., vol. 52, no. 6, pp. 23892402, 2013.
- [4] [Abdar et al., 2020] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, and R. Gururajan, "A new nested ensemble technique for automated diagnosis of breast cancer," Pattern Recognit. Lett., vol. 132, pp. 123131, Apr. 2020.
- [5] [Kok et al., 2020] C. Kok, V. Jahmunah, S. L. Oh, X. Zhou, R. Gururajan, X. Tao, K. H. Cheong, R. Gururajan, F. Molinari, and U. R. Acharya, "Automated prediction of sepsis using temporal convolutional network," Comput. Biol. Med., vol. 127, Dec. 2020, Art. no. 103957.
- [6] [Bargshady et al., 2019] G. Bargshady, J. Soar, X. Zhou, R. C. Deo, F. Whittaker, and H. Wang, "A joint deep neural network model for pain recognition from face," in Proc. IEEE 4th Int. Conf. Comput. Commun. Syst. (ICCCS), Feb. 2019, pp. 5256.
- [7] [Bargshady et al., 2020] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Ensemble neural network approach detecting pain intensity from facial expressions," Artif. Intell. Med., vol. 109, Sep. 2020, Art. no. 101954.
- [8] [Nasarian et al., 2020] E. Nasarian, M. Abdar, M. A. Fahami, R. Alizadehsani, S. Hussain, M. E. Basiri, M. Zomorodi-Moghadam, X. Zhou, P. Pawiak, U. R. Acharya, R.-S. Tan, and N. Sarrafzadegan, "Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach," Pattern Recognit. Lett., vol. 133, pp. 3340, May 2020.
- [9] [Zomorodi-Moghadam et al., 2021] M. Zomorodi-Moghadam, M. Abdar, Z. Davarzani, X. Zhou, P. Pawiak, and U. R. Acharya, "Hybrid particle swarm optimization for rule discovery in the diagnosis of coronary artery disease," Expert Syst., vol. 38, no. 1, Jan. 2021, Art. no. e12485.
- [10] [Abdar et al., 2019] M. Abdar, E. Nasarian, X. Zhou, G. Bargshady, V. N. Wijayaningrum, and S. Hussain, "Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach," in Proc. IEEE 4th Int. Conf. Comput. Commun. Syst. (ICCCS), Feb. 2019, pp. 2630.
- [11] [Albawi et al., 2017] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in Proc. Int. Conf. Eng. Technol. (ICET), Aug. 2017, pp. 16.
- [12] [Abiodun et al., 2018] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in articial neural network applications: A survey," Heliyon, vol. 4, no. 11, Nov. 2018, Art. no. e00938.
- [13] [Jin et al., 2017] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Trans. Image Process., vol. 26, no. 9, pp. 45094522, Sep. 2017.
- [14] [Zhang et al., 2020] Qingguo Zhang, Bizhen Lian, Ping Cao, Yong Sang, Wanli Huang, and Lianyong Qi, "Multi-Source Medical Data Integration and Mining for Healthcare Services," IEEE Access, VOLUME 8, pp. 165010- 165017, Sep 2020.

## Cloud Computing: Types, Security Issues, Benefits

**Miss. Rutuja Naresh Turkhade.**

M.SC. Ilyr (CS)  
Department of Computer Science  
Vidya Bharati Mahavidyalaya,  
Amravati.  
rutujaturkhade@gmail.com

**Miss. Nirja Vinod Deshmukh.**

M.SC. Ilyr (CS)  
Department of Computer Science  
Vidya Bharati Mahavidyalaya,  
Amravati.  
nirjadeshmukh3538@gmail.com

**Dr. Shilpa B. Sarvaiya.**

Department of Computer Science  
Vidya Bharati Mahavidyalaya,  
Amravati.  
Sarvaiya.shilpa@gmail.com

### Abstract:-

The global competition have extended the scope of trade and extended the supply chains causing the information management to be an essential part of the businesses. Many new technologies have recently been adapted by organizations in the era of digitalization. Cloud computing had an important place in these technologies and has been integrated widely to organizations. In recent days unexpected and risky periods such as the global pandemic increased the interest towards cloud computing topic in both academically and practically. The purpose of this study is to analyze and classify the contributions of the studies published in cloud computing field. This research summarizes the current research attempts, discovers the research gaps and provides a research agenda for the future research on cloud computing within the context of inforAbstract. Computing increases the flexibility and access of educational users to a wide range of educational resources. This includes access to infrastructure, software, hardware, and platform at any time in any place provided there is internet access.

### 1. Introduction:-

The influence of cloud computing on business and end users is impossible to overstate: the ubiquitous presence of software that operates on cloud networks has altered many elements of daily life. Start-ups and organisations can save costs and expand their offerings by utilising cloud computing instead of purchasing and managing all of the necessary hardware and software[1][2]. Independent developers now have the ability to create apps and internet services that are available worldwide. Researchers can now share and analyse data at scales previously only available to large-scale operations. Furthermore, internet users may instantly access software and storage to produce, exchange, and store digital media in quantities much exceeding their personal computing power. Despite the fact that cloud computing is becoming more prevalent, many people are unaware of its specifics. little research has been devoted to the continuance use in an organizational setting. The collective nature of all these entities is known as the Cloud. While respondents from all types and sizes of institutions showed some similarities in their perceptions of the cloud and cloud usage, there are some notable differences that are highlighted in the key findings. This paper also aims to identify the benefits and limitations of SaaS in higher educational institutions, closes with a discussion of the research limitations, contribution, and future directions.

### 2. CLOUD COMPUTING DEFINITION

The term "cloud" refers to software tools and services that run over the Internet or through a web browser. This is in contrast to traditional systems, which are limited to running on a single machine. As a result, cloud computing refers to the delivery of services such as data storage, networking, and servers over the Internet. Users can save files on a distant database rather than a hard drive or storage tool with cloud computing. A device that is linked to the internet is all that is required to access the database.

## 2.1 What is cloud computing?



Cloud computing is on-demand access, via the internet, to computing resources—applications, servers (physical servers and virtual servers), data storage, development tools, networking capabilities, and more—hosted at a remote data center managed by a cloud services provider (or CSP). The CSP makes these resources available for a monthly subscription fee or bills them according to usage.

Compared to traditional on-premises IT, and depending on the cloud services you select, cloud computing helps do the following:

- **Lower IT costs:** Cloud lets you offload some or most of the costs and effort of purchasing, installing, configuring, and managing your own on-premises infrastructure.
- **Improve agility and time-to-value:** With cloud, your organization can start using enterprise applications in minutes, instead of waiting weeks or months for IT to respond to a request, purchase and configure supporting hardware, and install software. Cloud also lets you empower certain users—specifically developers and data scientists—to help themselves to software and support infrastructure.
- **Scale more easily and cost-effectively:** Cloud provides elasticity—instead of purchasing excess capacity that sits unused during slow periods, you can scale capacity up and down in response to spikes and dips in traffic. You can also take advantage of your cloud provider’s global network to spread your applications closer to users around the world.

The term ‘cloud computing’ also refers to the technology that makes cloud work. This includes some form of *virtualized IT infrastructure*—servers, operating system software, networking, and other infrastructure that’s abstracted, using special software, so that it can be pooled and divided irrespective of physical hardware boundaries. For example, a single hardware server can be divided into multiple virtual servers.

Virtualization enables cloud providers to make maximum use of their data center resources.

Not surprisingly, many corporations have adopted the cloud delivery model for their on-premises infrastructure so they can realize maximum utilization and cost savings vs. traditional IT infrastructure and offer the same self-service and agility to their end-users.

If you use a computer or mobile device at home or at work, you almost certainly use some form of cloud computing every day, whether it’s a cloud application like Google Gmail or Salesforce, streaming media like Netflix, or cloud file storage like Dropbox.

## 2.2 History of Cloud Computing :-

In this, we will discuss the history of Cloud computing. Before Computing was come into existence, client Server Architecture was used where all the data and control of client resides in Server side. If a single user want to access some data, firstly user need to connect to the

server and after that user will get appropriate access. But it has many disadvantages. So, After Client Server computing, Distributed Computing was come into existence, in this type of computing all computers are networked together with the help of this, user can share their resources when needed. It also has certain limitations. So in order to remove limitations faced in distributed system, cloud computing was emerged.

- During 1961, John MacCharly delivered his speech at MIT that “Computing Can be sold as a Utility, like Water and Electricity.” According to John MacCharly it was a brilliant idea. But people at that time don’t want to adopt this technology. They thought the technology they are using efficient enough for them. So, this concept of computing was not appreciated much so and very less will research on it. But as the time fleet the technology caught the idea after few years this idea is implemented. So, this is implemented by Salesforce.com in 1999.
- This company started delivering an enterprise application over the internet and this way the boom of Cloud Computing was started.
- In 2002, Amazon started Amazon Web Services (AWS), Amazon will provide storage, computation over the internet. In 2006 Amazon will launch Elastic Compute Cloud Commercial Service which is open for Everybody touse.
- After that in 2009, Google Play also started providing Cloud Computing Enterprise Application as other companies will see the emergence of cloud Computing they also started providing their cloud services. Thus, in 2009, Microsoft launch Microsoft Azure and after that other companies like Alibaba, IBM, Oracle, HP also introduces their Cloud Services. In today the Cloud Computing become very popular and important skill.

### **2.3 TYPES OF CLOUD COMPUTING**

**Public Cloud:** The public cloud is a computing service supplied by the third party providers atop the public internet . These services are available for any user who wants to use them and they have to pay only for the services they consumed.

**Private Cloud:** The computing services provided over the internet or private network come under the private cloud and these services are offered only to the selected users in place of common people . A higher security and privacy is delegated by private clouds through the firewall and internal hosting .

**Hybrid Cloud:** Hybrid cloud is the combination of public cloud and private cloud. In the hybrid cloud, each cloud can be managed independently but data and applications can be shared among the clouds in the hybrid cloud(19) .

### **2.4 BENEFITS OF CLOUD COMPUTING**

**Cost Saving:** In cloud computing users have to only pay for the services they consumed.

**Maintenance cost is low** as user do not need to purchase the infrastructure **Flexibility:** Cloud computing is scalable. The rapid scale up and down in the operations of your business may require quick adjustment of hardware and resources so in order to manage this variations cloud computing provide flexibility. **Enhanced Security:** Cloud computing provide high security by using the data encryption, strong access controls, key management, and security intelligence.

### **2.5 SECURITY ISSUES IN CLOUD COMPUTING**

There are various security issues for cloud computing as it comprises of numerous advancements including systems, databases, working frameworks, virtualization, asset planning, exchange administration, stack adjusting, simultaneousness control and memory administration. Similarly, security issues for greater number of these frameworks and technology are pertinent to Cloud computing. According to the RSA conference which was

---

conducted in the March 2016, the CSA (Cloud Security Alliance) has released the list known as Treacherous 12, which includes the top 12 Cloud Computing threats in 2016. The following are the 12 threats in cloud computing .

1 Data Breaches

2 Compromised credentials and broken authentication

3. Hacked Interfaces and APIs

4 Exploited system vulnerabilities

5 Account Hijacking

**1. Data Breaches:** Due to the improved technology, large amount of data is stored in cloud servers, which becomes a target for the hackers. More the amount of data exposed, greater will be the damage to the society and users. The exposure of personal profile would be a normal one, but breaches which involve health information, trading secrets, intellectual property rights would bring a larger destruction. Though Cloud provider typically disposed security controls to protect their environments, it is enterprises which are responsible for securing their own data in cloud. Use of multi-factor authentication and encoding the data or information so that only authorized users can access it.

**2. Compromised credential sand broken authentication:** Data breaches and other attacks frequently result from slack authentications, weak passwords, poor key or certificate management. Sometimes, not only organizations even we forget to remove the access after our job is done. We can consider for example, the Gmail account if we login in the public accessing places (internet cafes) and forget to logout after our use, exposes our own private data to others. It is our responsibility to remember everything and take care. To avoid these issues, Multi-factor authentications such as one-time passwords, phone-based authentications, OTPs, security questions would make the attacker harder to login from stolen passwords. The rotation of cryptographic keys periodically will not only keep the records secure but also make the resources difficult for the attackers who use keys without authorization.

**3. Hacked Interfaces and APIs:** At present, every cloud service provides APIs. They are used to manage the cloud services, management, orchestration, monitoring. The interfaces and APIs which are weak would expose the authorizations to security issues like confidentiality, integrity, availability and accountability. It is recommended by CSA, to focus on threat modeling applications such as architecture/ design which are the primary concepts for the future developments and also to examine the flaws in the security-coding reviews and high level of testing.

**4. Exploited system vulnerabilities:** We have been facing the problem of bugs since a very longtime. One can say that they are always observed in one or the form. As the usage of technology has increased in a wide range, these vulnerabilities had become a bigger issue. The sharing of memory, data bases and other data among the organizations would lead to data crash or reports larger bugs and later on even may be affected by virus too. To eschew these bugs and system vulnerabilities one may probably have to scan the systems, mobile phones etc. regularly and try to find the solutions for the reported bugs.

**5. Account Hijacking:** Security concerns with cloud computing . One of the most common and daily heard issues in the society at present(12)(13).

## **2.6 CHARACTERISTICS OF CLOUD COMPUTING**

There has been much discussion in industry and academia about what cloud computing actually means . The US National Institute of Standards and Technology (NIST) has developed a working definition that covers the commonly agreed aspects of cloud computing . It summarizes cloud computing as: “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal



management effort or service provider interaction". According to this definition, cloud computing has the five essential characteristics, 1) On-demand self-service. 2) Broad network access. 3) Resource pooling. 4) Rapid elasticity. 5) Measured Service. Cloud computing is an emerging distributed computing paradigm that promises to offer cost-effective scalable on demand services to users, without the need for large up-front infrastructure investments . One of the main reasons for the success of cloud computing is the role it has played in eliminating the size of an enterprise as a critical factor in its economic success. An excellent example of this change is the notion of data centers which eliminate the need for small companies to make a large capital expenditure in building an infrastructure to create a global customer base

#### **Conclusion:-**

Cloud computing will affect large part of computer industry including Software companies, Internet service providers. Cloud computing makes it very easy for companies to provide their products to end-user without worrying about hardware configurations and other requirements of servers. The cloud computing and virtualization are distinguished by the fact that all of the control plane activities that center around creation, management, and maintenance of the virtual environment, are outsourced to an automated layer that is called as an API and other management servers for the cloud management.

In simple words, the virtualization is a part of cloud computing where manual management is done for interacting with a hypervisor. On the other hand, in cloud computing, the activities are self-managing where an API (Application Program Interface) is used so that the users can self-consume the cloud service.

#### **References :-**

- [1] L.Wang, J. Tao, M. Kunze, A. C. Castellanos, D. Kramer, and W. Karl, "Scientific cloud computing: Early definition and experience," 2008 10<sup>th</sup> IEEE International Conference on High Performance Computing and communications, 2008.
- [2] <https://www.ugc.edu.hk>.
- [3] <https://www.academia.edu>.
- [4] veritisadmn,"cloud computing trends,benefits,"go to veritis group inc.,12-oct- 2022.[online].
- [5] l qian,z.luo,y.du,and l.guo,"cloud computing:anoverview,"incloudcom '09:proceedings of the 1<sup>st</sup> international conference on cloud computing.springer-verlag pp.626-631,2009.
- [6] g.lin,d.fu,j.zhu,and g.dasmalchi,"cloud computing:IT as a service,"IT professional,vol.11,no.2,pp. 10-13,2009.
- [7] L.gonzalez, L.merino,j.caccres, and M. Lindner, " A Break in the Clouds: Towards a Cloud Definition", Computer Communication Review,39(1),2009.
- [8] D.Plummer, T. Bittman, T. Austin, D. Cearly, and D.Smith, "Cloud computing: Defining and describing an emerging phenomenon", Technical report, Gartner, 2008.
- [9] J. Staten, S. Yates, F. Gillett, W. Saleh, and R. Dines, "Is cloud computing ready for the enterprise?", Technical report, Forrester Research, March 2008.
- [10] Herhalt, J., Cochrane, K..Exploring the Cloud: A Global Study of Governments' Adoption of Cloud(2012).
- [11] <http://www.appcore.com/types-cloud-computing-private-public-hybrid-clouds/>
- [12] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, and G. S. Salvador, "A Cloud Computing Solution for Patient ' s Data Collection in Health Care Institutions," no. ii, pp. 95–99, 2010.
- [13] J. S. Sengar, "SURVEY : Reputation and Trust Management in VANETs," Int. J. Grid Distrib. Comput., vol. 8, no. 4, pp. 301–306, 2015.

## Pre-Processing techniques Using Weka Tool

**Priya Janardhan Deshmukh**

Department of Research and PG Studies in Science & Management  
M.C.A. PROGRAMME, Vidya Bharati Mahavidyalaya, Amravati

### ABSTRACT

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task. This paper is a review paper that introduces the key principle of data preprocessing technique on Training Dataset Table using WEKA tool such as Add, Remove, Normalization technique. "Weska" is a data mining tool. In this paper we are describing the steps of how to use preprocessing technique on training Dataset using Weka Tool.

### General Terms

The paper contain some general terms as Supervised, Unsupervised, Best First Search Algorithms, etc.

### Keywords

Data mining, WEKA tool, Data pre-processing, Data set, Add, Remove, Normalization, Knowledge Flow

## 1. INTRODUCTION

Data mining is a disciplinary sub domain of computer science. Data mining has been defined as the implicit extraction, prior unknown and potentially useful information from historical data databases. It uses machine learning, statistical techniques to discover and represent knowledge in a form, which is simply comprehensive to humans. A number of algorithms have been developed to extract discover knowledge patterns.

Data mining is there search step of the KDD (knowledge discovery in databases process). Data pre-processing, clustering, classification is the popular technologies in data mining. Data mining tools predict behaviors and upcoming trends, helps businesses to make change in knowledge-driven decisions. Data mining tools can solution, business queries that traditionally were too time consuming to resolve.

In this paper the main focus is to detail the ability of the data pre-processing & Weka background.

## 2. DATA PRE-PROCESSING

### 2.1 Why to process the Data?

If there is much tangential & redundant information available, noisy & unreliable data and Errors in transmission of data and instruments that collect the faulty data are present then knowledge discovery in Database (KDD) during the training phase is more

difficult. Data filtering & preparation can take considerable amount of time. Data pre-processing includes cleaning is the final training set.[1]

### 2.2 Why Data Pre-processing is & its Methods

Raw data is highly susceptible to noise, lacking attributes values & inconsistency occurred from different data sources. This quality of data infect the DM results. In order to enhance the efficiency & to improve the quality of data & accordingly, the mining results of raw data is pre-processed. Quality decisions must be based on the quality data. **By data processing, standard quality of data can be maintained measured in term of accuracy, completeness, consistency, timeliness, interpretability,**

**believability.**

Duplicates records also need data cleaning.

Data Pre-processing & transformation of the initial dataset. The process of Data Pre-processing are described below:

+**Data Cleaning**:- fill in missing values, resolve inconsistencies & smooth noisy data.

+**Data Integration**:-using multiple databases, or files.

+**Data Transformation**:-aggregation and normalization.

+**Data Reduction**:-reducing the volume but producing similar analytical results.[2]

**2.3 Data Pre-Processing Method**

Fig 1: The Steps of Data Pre- Processing Technologies or Method

**3. ADD Technique Using Weka Tool**

Unsupervised-Attribute-Add.

It is an instance filter that adds a new attribute to the dataset.

The new attribute will contain all missing values.

**CAPABILITIES**

Class—Binary class, Date class, Empty nominal class, Missing values, No class, Nominal class, Numeric class, Relational class, String class, Unary class.

Attributes—Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Interfaces—StreamableFilter, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstanceHandler

Additional

Minimum number of instances:0

**4.Remove Technique Using Weka Tool**

UnSupervised-Attribute-Remove

It is a filter that removes a range of attributes from the dataset.

Will re-order the remaining attributes if invert matching sense is turned on and the attribute column indices are not specified in ascending order.

**CAPABILITIES**

Class—Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class

Attributes—Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Interfaces—StreamableFilter, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstanceHandler

Additional

Minimum number of instances:0

## 5. Normalize Technique Using Weka Tool

Unsupervised-Attribute-Normalize

It normalizes all numeric values in the given dataset (apart from the class attribute, if set)

By default, the resulting values are in [0,1] for the data used to compute the normalization intervals. But with the scale and translation parameters one can change that, e.g. with scale = 2.0 and the translation = -1.0 you get values in the range [-1,1].

### CAPABILITIES

Class—Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class

Attributes—Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Interfaces—Sourceable, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstancesHandler

Additional

Minimum number of instances:0

### 5.1 Min-Max Normalization

To transform data from one range of [min,max] to other range [new\_min.new\_max }

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$$

Here new\_max=1 and new\_min=0

### 5.2 Normalization Benefits:

Data normalization can help avoid data quality issues, reduce data redundancy, improve data analysis, and enhance data security. It can eliminate errors, inconsistencies, duplicates, or missing values that can affect the accuracy of your data and analysis.

### 5.3 Normalize Data Using Knowledge Flow

In this technique we use, Best First Search Technique.

How Best First Search in normalization

The **best first search in artificial intelligence** is an informed search that utilizes an evaluation function to opt for the promising node among the numerous available nodes

before switching (transverse) to the next node. The **best first search algorithm in AI** utilizes two lists of monitoring the transversal while searching for graph space, i.e., Open and CLOSED list. An Open list monitors the immediate nodes available to transverse at the moment. In contrast, the CLOSED list monitors the nodes that are being transferred already.

The algorithm works by evaluating the cost of each possible path and then expanding the path with the lowest cost. This process is repeated until the goal is reached. Greedy Best-First Search has several advantages, including being simple and easy to implement, fast and efficient, and having low memory requirements.

Best first search is an instance of graph search algorithm in which a node is selected for expansion based on evaluation function  $f(n)$ .

## Experimental Results

Weather Table after adding new attribute CLIMATE:

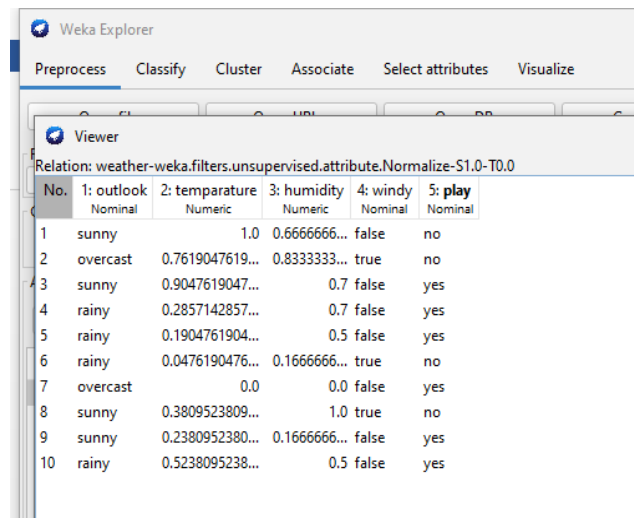
Viewer						
relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-NClimate-L-Clast-W1.0						
No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal	6: Climate Nominal
1	sunny	85.0	85.0	false	no	
2	overcast	80.0	90.0	true	no	
3	sunny	83.0	86.0	false	yes	
4	rainy	70.0	86.0	false	yes	
5	rainy	68.0	80.0	false	yes	
6	rainy	65.0	70.0	true	no	
7	overcast	64.0	65.0	false	yes	
8	sunny	72.0	95.0	true	no	
9	sunny	69.0	70.0	false	yes	
10	rainy	75.0	80.0	false	yes	

### MIN-MAX NORMALIZATION

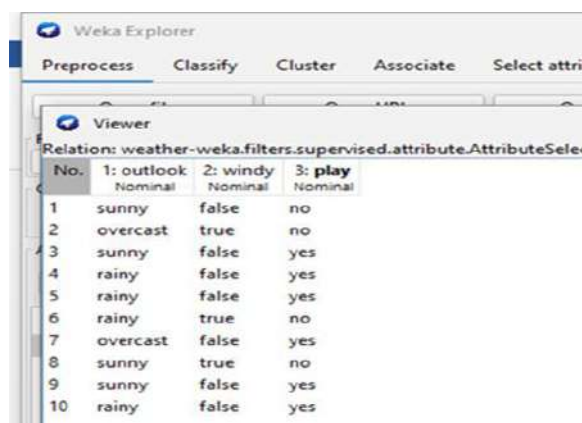
Weather Table after removing attribute windy, play:

relation: weather-weka.filters.unsupervised.attribute.Add-TNO				
No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: Climate Nominal
1	sunny	85.0	85.0	
2	overcast	80.0	90.0	
3	sunny	83.0	86.0	
4	rainy	70.0	86.0	
5	rainy	68.0	80.0	
6	rainy	65.0	70.0	
7	overcast	64.0	65.0	
8	sunny	72.0	95.0	
9	sunny	69.0	70.0	
10	rainy	75.0	80.0	

## NORMALIZATION USING KNOWLEDGE FLOW



No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	1.0	0.6666666...	false	no
2	overcast	0.7619047619...	0.8333333...	true	no
3	sunny	0.9047619047...	0.7	false	yes
4	rainy	0.2857142857...	0.7	false	yes
5	rainy	0.1904761904...	0.5	false	yes
6	rainy	0.0476190476...	0.1666666...	true	no
7	overcast	0.0	0.0	false	yes
8	sunny	0.3809523809...	1.0	true	no
9	sunny	0.2380952380...	0.1666666...	false	yes
10	rainy	0.5238095238...	0.5	false	yes



No.	1: outlook Nominal	2: windy Nominal	3: play Nominal
1	sunny	false	no
2	overcast	true	no
3	sunny	false	yes
4	rainy	false	yes
5	rainy	false	yes
6	rainy	true	no
7	overcast	false	yes
8	sunny	true	no
9	sunny	false	yes
10	rainy	false	yes

## 6. CONCLUSION

Conclude that data preprocessing techniques have an efficient, effective and important role in preparation, analysis, process large *data*-scale

## 7. REFERENCES

1. <https://youtu.be/chgLILDwuDo?si=SjntEKr5sEAZPr66>
2. <https://youtu.be/RIoUO1ndFNk?si=nqM1ycUIRfN194T1>
3. <https://www.studocu.com/in/document/gandhi-institute-of->

## Web Data Mining: A Comprehensive Analysis of Types, Tools, and Techniques

**Ms. Anjali V. Parasmode**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
[parasmodeanjali7@gmail.com](mailto:parasmodeanjali7@gmail.com)

**Dr. Sonali R. Chavan**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
[sonal.chavan3989@gmail.com](mailto:sonal.chavan3989@gmail.com)

### **Abstract-**

Web mining plays a pivotal role in extracting valuable insights from the vast ocean of information available on the World Wide Web. This research paper delves into the multifaceted landscape of web mining, providing a comprehensive analysis of its three distinct types: web content mining, web structure mining, and web usage mining. The study explores the underlying principles and methodologies associated with each type, shedding light on the intricate processes involved in uncovering hidden patterns and knowledge from web data. From traditional algorithms to emerging technologies, the analysis encompasses a wide spectrum of approaches that researchers and practitioners can leverage for effective web mining.

*Keywords-* Web mining, Content mining, structure mining, usage mining

### **I. INTRODUCTION**

Web data mining is the application of web data mining technique which is an unstructured or semi structured data and it inevitably locates and extracts potentially useful and previously unknown information or knowledge from the web[1]. Huge amount of information available on the world wide web leads to the mining of Web. So, web mining can be defined as the use of the data mining techniques to automatically discover and extract information from web documents and services[2].

Data mining is a process of analysing usable information and extract data from large data warehouses involving different patterns, intelligent methods, algorithm and tools. This process can help business to analyse data user behaviour and predict future trends. Web is one of the type of techniques use in data mining. The main purpose of web mining is to automatically extract information from the web. For discovering useful data (videos, tables, audio, images etc.) from the web different techniques and tools are used. Information over the internet is huge and increasing with passage to time due to which size of data bases are also growing. Digging knowledgeable information and analysing the data sets for the relevant data is much difficult because data over the internet in not in plain text. It could be unstructured data, multimedia, table, tag [3]. *Purpose of this paper is to describe web mining its three different types, tools and techniques. All three types are explained in detailed and main focus is on web usage mining, its technique.*

### **II. LITERATURE REVIEW**

Data mining is a process of discovering knowledge from data warehouse. This knowledge can be classified in different rules and patterns that can help user/organization to analyse collective data and predicted decision processes [4].

Centralized database of any organization is known as Data warehouse, where all data is stored in a single huge database. Data mining is a method that is used by organization to get useful

information from raw data. Software's are implemented to look for needed patterns in huge amount of data(data warehouse) that can help business to learn about their customers, predict behaviour and improve marketing strategies.

Web mining is actually an area of data mining related to the information available on internet. It is a concept of extracting informative data available on web pages over the internet. Users use different search engines to fetch their required data from the internet, that informative and user needed is data is discovered through mining technique called Web Mining. Different tools and algorithms are used for extraction of data from web pages that includes web documents, images etc. Web mining is rapidly becoming very important due to size of text documents increasing over the internet and finding relevant patterns, knowledge and informative data is very hard and time consuming if it is done manually. Structure (Hyperlinks), Usage (visited pages, data use), content (text document, pages) are included in information gathered through Web mining [5].

Term World Wide Web is related to the combination of web documents, videos, audios etc. Some processes included in web mining are: Information Retrieval is a process of retrieving relevant and useful information over the web. Information retrieval has more focuses on selection of relevant data from large collection of databases and discovering new knowledge from large quantity of data to response user query. IR steps include searching, filtering and matching. Information extraction is an automatic process of extracting analysed data (structured). IE is a task that work same like information retrieval but more focuses on extracting relevant facts[6].

### III. CATEGORIES OF WEB MINING

Web Mining is sub categorized into three types

#### A. Web Content Mining

Content Mining is a process of Web Mining in which needful informative data is extracted from web sites (WWW). Content includes audio, video, text documents, hyperlinks and structured Web contents are designed to deliver data to users in the form of text, list, images, videos and tables. Over last few decades the amount of web pages (HTML) increases to billions and still continues to grow. Searching query into billions of web documents is very difficult and time-consuming task, content mining extracts queried data by performing different mining techniques and narrow down the search data which become easy to find required user data [7].

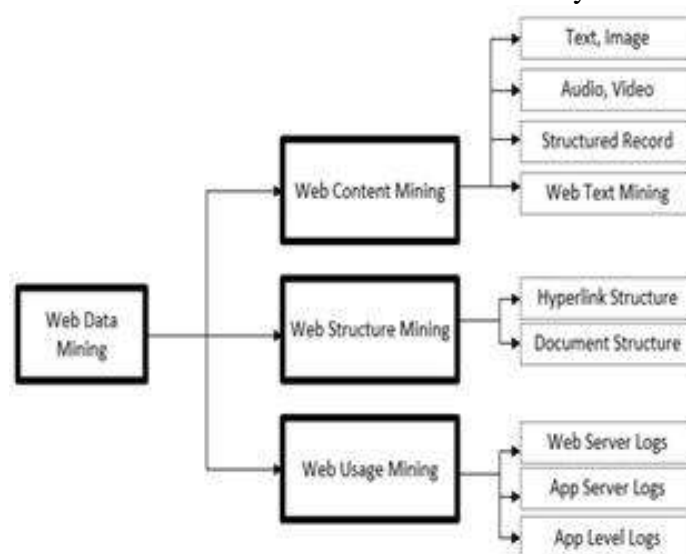


Fig 1. Web Data Mining



### B. Web Structure Mining

Web mining techniques are very useful to discover knowledgeable data from web. Structure mining is one of the core techniques of web mining which deals with hyperlinks structure. Structure mining basically shows the structured summary of the website. It identifies relationship between linked web pages of websites. Continuous growth of data over the internet become a challenging task to find informative and required data [8].

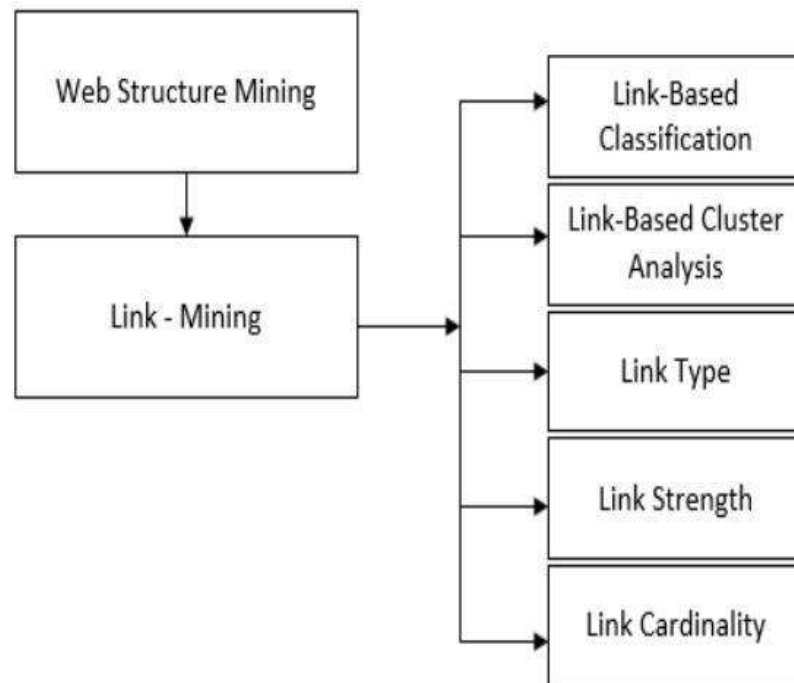


Fig. 2 Structure of Web Mining

### C. Web Usage Mining

Web usage mining also called log mining is a process of recording user access data on the web and collect data in form of logs. After visiting any website user leaves some information behind such as visiting time, IP address, visited pages etc. This information is collected, analyzed and store in logs. Which helps to understand user behavior and later can improves website structure [9].

Web usage mining dig and analyze data present in log files which contains user access patterns. Main purpose of web usage mining is to observer user behavior at the time of his interacting with web. There are two types of patterns tracking i.e. general tracking and customized tracking. In general tracking information is collected from web page history. In customized tracking the information is gathered for specific user[10].

## IV. CONCLUSION

Web Data mining is a concept that helps to find information which is needed from large data warehouses by using different techniques. It is also used to analyze past data and improve future strategies. Web data mining is considered as sub approach of data mining that focuses on gathering information from web. Web is a large domain that contains data in various forms i.e.: images, tables, text, videos, etc. In this paper we described three important types of web data mining that can help in finding informative data. Web content mining is useful in terms of exploring data from text, table, images etc. Web structure mining classifies relationships between linked web pages. Web usage mining is also an important type that stores user access data and get information about specific user from logs.

---

**REFERENCES**

- [1] C. C. J. Hryhoruk and C. K. Leung, "Web Mining from Interpretable Compressed Representation of Sparse Web," 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Niagara Falls, ON, Canada, 2022, pp. 620-627, doi: 10.1109/WI-IAT55865.2022.00097.
- [2] K. Jayamalini and M. Ponnavaikko, "Research on web data mining concepts, techniques and applications," 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), Chennai, India, 2017, pp. 1-5, doi: 10.1109/ICAMMAET.2017.8186676.
- [3] R. K. Shukla, P. Sharma, N. Samaiya and M. Kherajani, "WEB USAGE MINING-A Study of Web data pattern detecting methodologies and its applications in Data Mining," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170690.
- [4] S. K. Malik and S. Rizvi, "Information Extraction Using Web Usage Mining, Web Scrapping and Semantic Annotation," 2011 International Conference on Computational Intelligence and Communication Networks, Gwalior, India, 2011, pp. 465-469, doi: 10.1109/CICN.2011.97.
- [5] S. P. Singh and Meenu, "Analysis of web site using web log expert tool based on web data mining," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8275961.
- [6] V. Rana and G. Singh, "Analysis of web mining technology and their impact on semantic web," 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), Ghaziabad, India, 2014, pp. 5-11, doi: 10.1109/CIPECH.2014.7019035
- [7] S. Yadao, A. V. Babu, M. Janarthanan and A. Bhaumik, "Web usage Mining: A Comparison of WUM Category Web Mining Algorithms," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 1020-1024, doi: 10.1109/ICICV50876.2021.9388539
- [8] Y. Li, "Research on Technology, Algorithm and Application of Web Mining," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, China, 2017, pp. 772-775, doi: 10.1109/CSE-EUC.2017.152.
- [9] V. Rana and G. Singh, "Analysis of web mining technology and their impact on semantic web," 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), Ghaziabad, India, 2014, pp. 5-11, doi: 10.1109/CIPECH.2014.7019035.
- [10] R. K. Shukla, P. Sharma, N. Samaiya and M. Kherajani, "WEB USAGE MINING-A Study of Web data pattern detecting methodologies and its applications in Data Mining," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170690.

## Feature Analysis of Fake News: Improving Fake News Detection on Social Media

<sup>1</sup>Pooja. G. Borkar

<sup>1</sup>Student Department of MCA, Vidya Bharti Mahavidyalaya, Amravati, India

### ABSTRACT

Fake news is a threat to society, and its spread can have real-world consequences in many situations. These days a lot of information is being shared over social media and we are not able to differentiate between which informatization is fake and which is real. People immediately start expressing their concern or sharing their opinion as soon as they come across a post, without verifying its authenticity. This further results in spreading of it. First, we extract features, such as visual sentiment, textual sentiment, behavioural reactions, and metadata, and then analyse various features for fake news prediction. We then run a machine learning experiment to classify posts that help improve fake news detection in social media.

**Keywords:** Fake News Detection, Social Media, Text Mining, Online social media, machine learning algorithm.

### Introduction

Nowadays, Internet has become an integral part of our life. It is not exaggerating to say that it has become the main part of our lifestyle. The role of print media e.g. newspaper and electronic media e.g. Facebook, Twitter, Instagram, YouTube, WhatsApp etc. The growth of social media platforms plays an important role in transformation of data.

The challenge in social media is gathering verified/authenticated news. Our review analyses how to detect the fake news on social media to overcome this problem.

Diffusion of fake news has become prevalent and poses a threat to society as well as to democracy as it can negatively influence the user's trust in governmental institutions. Major events such as the 2016 United States presidential election have involved fake news that proliferate on social media and influence voters. More recently, in the COVID-19 pandemic, an "infodemic" of information that contradicts official advice related to vaccines and health measures have rapidly proliferated on social media, which caused people to ignore recommended health guidelines and ultimately threaten public health. False rumours have also played a role in the reaction in other situations, such as terrorist attacks, by causing false information to propagate among the public and drown out contradicting facts.

On social media the spread of fake news is a significant issue, and fake news spreads faster than real news, possibly due to the emotional reactions of readers and the novelty of fake news. Fake news can be weaponized by attackers to influence online user opinion, such as during the 2016 US and 2017 French election. The major

part of spread of fake news is due to human behaviour, there is a clear need to investigate and detect the emo-

tional and behavioural factors in the fake news that drive humans to share fake posts.

Therefore, we aim to conduct a study of different categories of features for fake news detection, namely emo-

tional data in images and texts, as well as metadata (e.g. the poster, the referenced URL) and behavioural data (e.g. post score, replay count). By doing so we will be able to comprehensively identify the characteristics of fake news that entice users and evaluate the importance of

different characteristics. We use machine learning(ML) techniques to classify fake news using the identified influential or impactful features. Lot of persons use social media not only to be in contact with their friends but also to gather news around us. Social media is very useful for news consumption. Flip side of this is, without substantiation false information also spread very fast over social media.

## • Literature Review

The focus of this review is to gather the numerous sequence of linked works done on the area of fake news

detection above the social network. Thereby, I proceed the survey from various areas like Facebook, Twitter

etc. with the intention of detecting the possible reliability of the sophisticated knowledge.

## A. Domain

In this review, will be working on the papers which have acquire the possible results in detecting the untruthful

news over Social media, Posts sharing attitude on Facebook, sentiment analysis on Facebook, identifying fake users and fake news in the Twitter social network, fraudulent attempt to obtain emotionable information on Facebook, analyzing fake content of Twitter, automatic real-time detection of malicious content on Facebook, Understanding users behavior on Facebook through opinion mining, the author presented the geometric deep learning approach for fake news detection were spread on Twitter, the author design the semi-supervised learning to detect fake news on Twitter. The author targeted political domain to identify the impostures on Facebook.

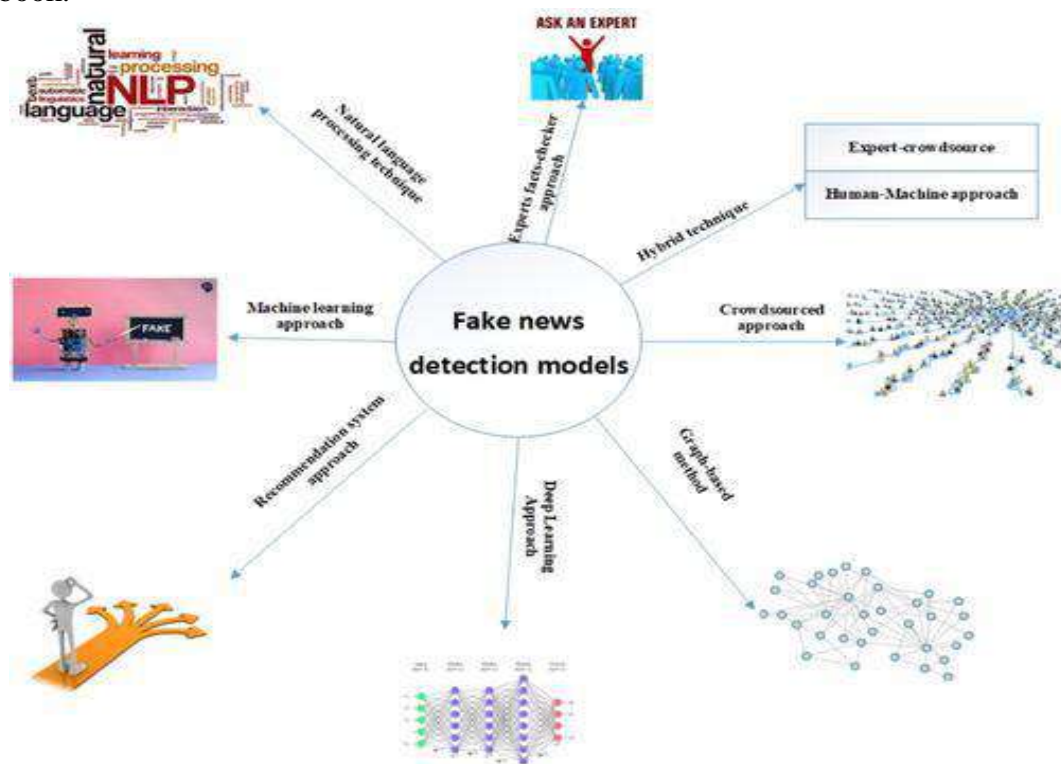


Fig 1. Fake news detection models.

## B. Methodology

The first step was to detect a credible clickbaits database, then compute the attributes and produce the data files for WEKA. That was not simple, therefore, we crawled the web to collect URLs for the clickbaits. We concentrated on social media websites that are likely to have more fake news or clickbait ads or articles, such as: Facebook, Forex and Reddit.

The second step, after gathering URLs in a file, a python script computed the attributes from the title and the content of the webpages. Finally, we pluck out the features from the web pages. The features are: keywords in Arabic and English, titles that starts with numbers, all caps words, carry question and exclamation marks, if user leave the page immediately, and content correlated to title.

Author design the algorithm to be good fit for checking the reliability of a news articles from Facebook using sentiment analysis, N-grams and Natural language Processing method to change the natural language to particular format. Using convolution Neural Network, the author make the geometric deep learning method of propagation-based approaches for fake news detection instead of using the content-based approaches. The author launch the set of features to measure the prediction performance and automatic fake news detection using the method lexical Features, Syntax Features, Semantic Features, Linguistic Features. The deep two path semi-supervised learning where one track is for supervised learning and a second is for unsupervised learning established by author for fake news detection using method Naive Bayes, Decision Tree, Adaboost, Support Vector Machine, bidirectional recurrent neural networks Twitter.

New review of literature highlights the methods Decision tree, Random forest, naïve Bayesian models used to found the false content on social media. Ahmed et al introduced the method to detect the false information spread above the social media using n-gram for feature extraction and machine learning techniques, namely, K-Nearest Neighbor, Support vector Machine, Logistic Regression, Linear Support Vector Machine, Decision tree, Stochastic Gradient Descent and uses Term Frequency-Inverted Document Frequency (TF-IDF) as feature extraction technique, and Linear Support Vector Machine (LSVM) as a classifier. Atodiresei et al form the method using linguistic approaches and naïve Bayes classifier to decide the identification of fake news and fake users on twitter. highlights linear regression and logistic regression to pre-determine future fake news and decide the fake news topics on social media.

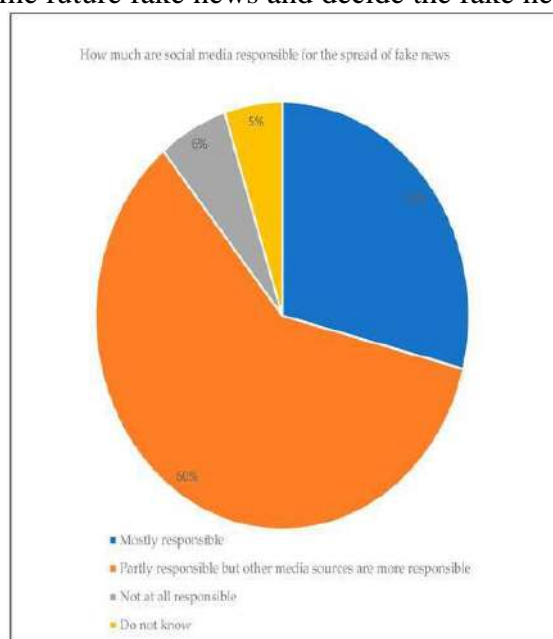


Fig 2. Social media responsible for spread of fake news

### c. Issues

The important issues showed in this articles is trouble in classification accuracy for confusing information on Facebook because of limitation in datasets and the length of the news articles. This article convey the issues like the model do not perform well on event specific data and do not claim that their dataset is representative of the entire Facebook community and also look challenges like gathering Facebook data and the amount of information. The limitation of work presents in this article is that the fake information or unsupported information are only from the unofficial news source not from the official newspapers.

### Conclusion

Fake news and Clickbaits interfere with the ability of a user to notice useful information from the Internet services mainly when news becomes critical for decision making. Considering the changing landscape of the modern business world , the issue of fake news has become more than just a marketing problem as it warrants serious efforts from security researchers. It is aggressive that any attempts to manipulate or troll the Internet through fake news or Clickbaits are countered with absolute effectiveness. We suggest a simple but effective approach to allow users install a simple tool into their personal browser and use it to detect and filter out potential Clickbaits. This paper presents the survey to fake news detection on social media, which is to recognize the community opinion to different posts of a user, and to identify the true news. Survey based on Fake news detection demonstrate using various machine Learning and Deep Learning Techniques. Machine Learning Algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Neural Network Models and Decision Trees are used to predetermine the future content and determine the inaccurate news and posts.

### References

1. Dong, X., Victor, U., Chowdhury, S., & Qian, L. (2019). Deep Two-path Semi-supervised Learning for Fake News Detection. arXiv preprint arXiv:1906.05659.
2. Lin, K. C., Wu, S. H., Chen, L. P., Ku, T., & Chen, G. D. (2014, August). Mining the user clusters on Facebook fan pages based on topic and sentiment analysis. In Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014) (pp. 627-632). IEEE.
3. Ahmed, A. A. A., Aljabouh, A., Donepudi, P. K., & Choi, M. S. (2021). Detecting Fake News using Machine Learning: A Systematic Literature Review. arXiv preprint arXiv:2102.04458. <https://arxiv.org/abs/2102.04458>
4. Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215-222. <https://doi.org/10.1016/j.procs.2018.10.171>
5. Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, GuangxuXun, KishlayJha, Lu Su, and Jing Gao.2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (London, United Kingdom) (KDD '18). Association for Computing Machinery, New York, NY, USA, 849–857. <https://doi.org/10.1145/3219819.3219903>
6. Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2021). A benchmark study on machine learning methods for fake news detection. *Computation and Language*. <https://arxiv.org/abs/1905.04749>
7. Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: A benchmark study for fake news detection. *Expert Systems with Applications*, 128, 201-213. <https://doi.org/10.1016/j.eswa.2019.03.036>
8. Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and VasudevaVarma. 2019. MVAE: Multimodal Variational Autoencoder for Fake News Detection. In The World Wide Web Conference (San Francisco, CA, USA) (WWW '19). Association for Computing Machinery, New York, NY, USA, 2915–2921. <https://doi.org/10.1145/3308558.331355>
9. Abdullah-All-Tanvir, Mahir, E. M., Akhter S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing &

- Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5, <https://doi.org/10.1109/ICSCC.2019.8843612>
10. Bahad, P., Saxena, P., & Kamal, R. (2019). Fake news detection using bi-directional LSTM- recurrent neural network. *Procedia Computer Science*, 165, 74-82. <https://doi.org/10.1016/j.procs.2020.01.072>
  11. Shu, K., Zhou, X., Wang, S., Zafarani, R., & Liu, H. (2019, August). The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 436 - 439). <https://doi.org/10.1145/3341161.3342927>
  12. Vishwakarma, D. K., Varshney, D., & Yadav, A. (2019). Detection and veracity analysis of fake news via scrapping and authenticating the web search. *Cognitive Systems Research*, 58, 217-229. <https://doi.org/10.1016/j.cogsys.2019.07.004>
  13. Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. (2019). SpotFake: A Multi-modal Framework for Fake News Detection. *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*. doi:10.1109/bigmm.2019.0044.
  14. Islam, M. S., Islam, M. A., Hossain, M. A., & Dey, J. J. (2016, December). Supervised approach of sentimentality extraction from Bengali facebook status. In *2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 383-387). IEEE.
  15. Tanwani, N., Kumar, S., Jalbani, A. H., Soomro, S., Channa, M. I., & Nizamani, Z. (2017, November). Student opinion mining regarding educational system using facebook group. In *2017 First International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT)* (pp. 1-5). IEEE.
  16. Mhamdi, C., Al-Emran, M., & Salloum, S. A. (2018). Text mining and analytics: A case study from news channels posts on Facebook. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 399-415). Springer, Cham.
  17. Toujani, R., & Akaichi, J. (2016, December). Fuzzy sentiment classification in social network Facebook's statuses mining. In *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)* (pp. 393-397). IEEE.
  18. Bozkır, A. S., Mazman, S. G., & Sezer, E. A. (2010, September). Identification of user patterns in social networks by data mining techniques: Facebook case. In *International Symposium on Information Management in a Changing World* (pp. 145-153). Springer, Berlin, Heidelberg.
  19. Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media—a survey. *Journal of Information and Telecommunication*, 5(2), 247-266. <https://doi.org/10.1080/24751839.2020.1847379>
  20. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in Social media with a BERT-based deep learning approach. *Multimedia tools and applications*, 80(8), 11765-11788. doi: 10.1007/s11042-020-10183-2
  21. Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting Fake News with Machine Learning Method. *2018 15<sup>th</sup> International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 528-531. <https://doi.org/10.1109/ECTICon.2018.8620051>
  22. Kesarwani, A., Chauhan, S. S., & Nair, A. R. (2020). Fake News Detection on Social Media using K-Nearest Neighbor Classifier. *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, Las Vegas, NV, USA, pp.1-4, <https://doi.org/10.1109/ICACCE49060.2020.9154997>
  23. Pratiwi, I. Y. R., Asmara, R. A., & Rahutomo, F. (2017). Study of hoax news detection using naïve bayes classifier in Indonesian language. *2017 11th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, pp.73-78. <https://doi.org/10.1109/ICTS.2017.8265649>
  24. Granik, M., & Mesyura, V. (2017). Fake news detection using naïve Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. doi:10.1109/ukrcon.2017.8100379.

## **Role of Internet of Things in Disease of Pate and Remote Monitoring System**

**<sup>1</sup>Rashmi A.Charjan, <sup>2</sup>Dr.Ajit Kumar, <sup>3</sup>Dr.G.K.Reddy**

<sup>1</sup>Shri Jagdishprasad Jhabarmal Tibrewala University, Vidya Nagari, Chudela, Churu- Jhunjhunu Road, Jhunjhunu, Rajasthan. Email- [rashmicharjan07@gmail.com](mailto:rashmicharjan07@gmail.com), Mob- 9766308150.

<sup>2</sup>Department Of Computer Science, JJT University, Jhunjhun, Rajasthan. E-Mail [ajit.kaswan@gmail.com](mailto:ajit.kaswan@gmail.com), Mob.No.9466535353.

<sup>3</sup>Associate Prof. & Head of Computer Science Department, Mahatma Fule Arts, Commerce & Sitaramji Chaudhari Science Mahavidyalaya, Warud.Dist: Amravati 444906, Maharashtra. Email- [reddygiridhar4@gmail.com](mailto:reddygiridhar4@gmail.com). Mob-9823288352

### **Abstract:**

This article explores the transformative role of the Internet of Things (IoT) in revolutionizing healthcare practices, specifically emphasizing its impact on disease management and remote patient monitoring. IoT technologies have paved the way for continuous and real-time patient monitoring, utilizing wearable devices and sensors to collect crucial health data. The integration of IoT in chronic disease management empowers patients and healthcare providers alike, fostering proactive interventions and personalized treatment plans. Remote patient engagement is enhanced through telehealth platforms, reducing the need for frequent hospital visits and facilitating seamless communication between patients and healthcare professionals. The vast amount of data generated by IoT devices allows for advanced analytics and predictive modeling, enabling healthcare providers to identify patterns, anticipate disease progression, and implement preventive measures. As technology continues to advance, the role of IoT in healthcare is set to expand, ushering in an era of connected and data-driven healthcare systems

**Keywords: Internet Of Things (Iot), Pate, Wearable Devices, Sensors, Patients, Healthcare, Monitoring, Telehealth.**

### **Introduction:**

The Internet of Things (IoT) has emerged as a groundbreaking technology with far-reaching implications across various industries. One of the most promising areas where IoT is making a significant impact is in the realm of healthcare, particularly in disease management and remote patient monitoring. This article explores the multifaceted role of IoT in revolutionizing healthcare, improving patient outcomes and transforming traditional healthcare practices especially on abnormality of head in Warud region.

In the rapidly evolving landscape of healthcare, technological innovations are playing a pivotal role in reshaping traditional practices and fostering a patient-centric approach. Among these innovations, the Internet of Things (IoT) has emerged as a transformative force, promising groundbreaking solutions to longstanding challenges. This article delves into the profound impact of IoT on disease of Pate management and remote patient monitoring, exploring how interconnected devices and advanced technologies are ushering in a new era of healthcare delivery.

### **Objectives:**

1. Analyze Remote Patient Engagement through Telehealth.
2. Investigate the Utilization of Data Analytics and Predictive Modeling.
3. Consider Future Trends and Developments
4. Examine the Impact on Patient Outcomes.



5. Assess Cost Efficiency and Resource Optimization.
6. Explore IoT's Contribution to Medication Adherence
7. Evaluate the Role of IoT in Chronic Disease Management.
8. Examine the Impact of IoT on Patient Monitoring.

**Challenges and Limitations**

The traditional healthcare paradigm, while delivering essential services, is not without its challenges and limitations. Understanding these issues is crucial for recognizing the need for transformative technologies like the Internet of Things (IoT) to reshape and enhance the healthcare landscape.

1. Episodic Nature of Care, Lack of Continuous Monitoring, Fragmented Data and Communication, Limited Patient Empowerment, Inefficiencies in Preventive Care, Challenges in Chronic Disease Management, Limited Utilization of Technology.

**Methodology**

Key characteristics of the current healthcare paradigm include: Episodic Care, Limited Continuous Monitoring, One-Size-Fits-All Approach, Hospital-Centric Focus, Reactive Management of Chronic Diseases, Limited Patient Engagement. IoT's Pioneering Role in Patient Monitoring, Wearable Devices and Sensors, Continuous and Real-Time Monitoring, Remote Patient Monitoring, Personalized Treatment Plans, Early Detection of Anomalies Patient Empowerment, Integration with Electronic Health Records (EHR).

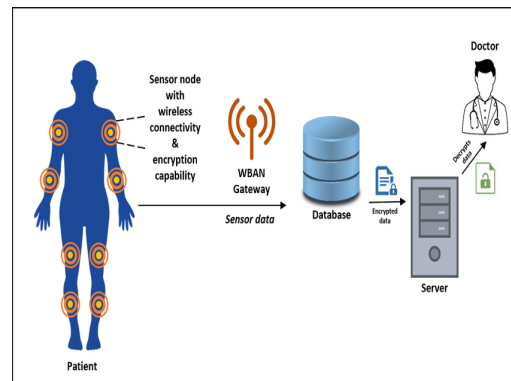
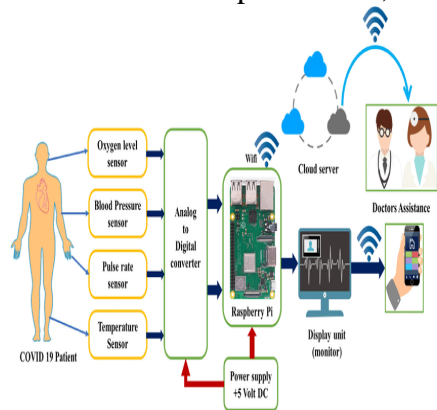


Figure 1. Block diagram of proposed IOT smart health system    Figure 2. Achitecture for patient monitoring system

**Work Observation:**

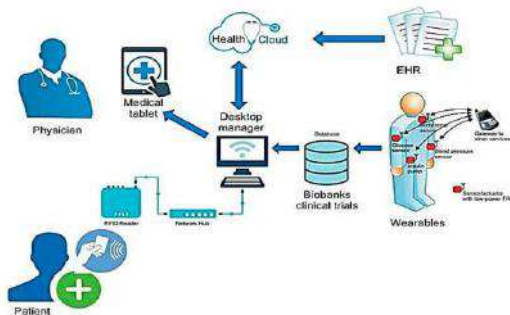


Figure 3. Revolutionary features of H-IoT in a hospital environment.

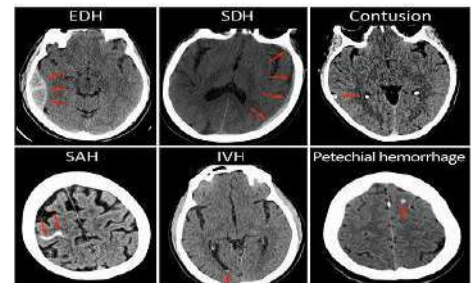


Figure 4. CT-Scan Report

---

In proposed research study, authors studied on abnormality of head in Warud region. Patients with a mean age of 41 and 66 percent of whom were male were assessed at two weeks and three, six and 12 months post-injury and found that patients in the SDH, SAH, and contusion groups failed to fully recover at 12 months post-injury and Effects ranged from mild to more severe disorders. It is found that approximately 9 percent of people develop intraventricular hemorrhage (IVH) or petechial hemorrhage in head-turning, sports, scooter and automobile accidents. Also 7 percent have an epidural hematoma (EDH) in which blood accumulates in the membrane covering the brain and skull, often seen in baseball sports injuries. The most common patterns of injury, affecting more than half of CT-positive patients, were a combination of subarachnoid hemorrhage (SAH), subdural hematoma (SDH), and contusion, which can result from injuries such as standing and falling.

Patients in the petechial hemorrhage and IVH hemorrhage groups tended to have more severe impairment falling into the low-moderate disability category. This level is seen to affect many areas of functioning such as social, leisure and employment activities up to 12 months after the injury. Patients with the EDH phenotype were then shown to perform significantly better at the six-month assessment. The research found that only 37 percent of people receive follow-up that includes simple interventions such as providing educational materials at discharge. This suggests that increased use of CT may expose patients to radiation and potentially increase the risk of cancer. CT is recommended only for patients with current suspected trauma and evidence of dementia and/or older age, physical trauma and severe headache.

#### **Future Enhancement:**

The role of the Internet of Things (IoT) in disease of Pate management and remote patient monitoring continues to evolve and several future enhancements are anticipated. These advancements aim to further improve healthcare outcomes, enhance patient experiences and streamline healthcare processes. Here are some potential future enhancements in the role of IoT in disease and remote patient monitoring:

Integration of Artificial Intelligence (AI). Predictive and Prescriptive Analytics. Edge Computing for Real-Time Processing. Enhanced Security and Privacy Measures. Wearable Device Innovations. Block chain for Data Integrity. Patient-Generated Health Data (PGHD) Integration. Enhanced User Interfaces and User Experience (UI/UX).

#### **Conclusion:**

In conclusion, this exploration of the transformative role of the Internet of Things (IoT) in healthcare underscores the profound impact. It is found that approximately 9 percent of people develop intraventricular hemorrhage (IVH) or petechial hemorrhage in head-turning, sports, scooter and automobile accidents. Also 7 percent have an epidural hematoma (EDH) in which blood accumulates in the membrane covering the brain and skull, often seen in baseball sports injuries it has on revolutionizing patient care, with a particular emphasis on disease management and remote patient monitoring. The integration of IoT technologies has ushered in a new era characterized by continuous and real-time monitoring, providing healthcare professionals with unprecedented insights into patients' health statuses.

---

**References**

1. Yogesh V, A. H Shanthakumara, "Smart Real-Time Health Monitoring Band Using Machine Learning and IoT", *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, pp.43-46, 2021.
2. Sapna Kumari C, Sahana R, Skanda K Mitta, Vinay K V, Yashas G S, "HEALTH MATE: Personal Health Tracker Based on the Internet of Things", *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pp.1-6, 2023.
3. R. Singh, Yasmeen, A. Srivastava, B. Bijeshdhyani and D. Deepa, "Machine learning based human interacted robotic intelligence to detect the different categories of speech," *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2023,
4. Gowtham S, Venkatesh L, Rajendra Varaprasad B, Diwakar SS, Aarthi N, "IoT Based Health Monitoring System", *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pp.1-6, 2021.
5. Kiran, Parameshahcari B D, Sunil Kumar D S, "Embedded System For Chronic Disease Patient Monitoring Using Iot", *2023 International Conference on Data Science and Network Security (ICDSNS)*, pp.1-5, 2023.
6. S.Stella Rani, T.Hemanth Pavan, A. Vanathi, B.S.Kiruthika Devi, "A Review on Patients Health Monitoring using IoT and Cloud Technologies", *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp.1-6, 2022.
7. K. Rajasekaran, M. Subash Chakaravarthi, P. Lokaswar, "Continuous Health Monitoring System for Patients Using IoT", *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol.1, pp.987-990, 2023.
8. Dr.G.K.Reddy and Miss.R.A.Charjan, Water Demand Registration System For Agriculture, *Ijrmets*, E-Issn-2582-5208, Volume03/Issue:12 /Decemer-2021, [www.Ijrmets.Com](http://www.Ijrmets.Com)
9. Miss.R.A.Charjan and Dr.G.K.Reddy, Android Application for Android Attendance System, *International Journal of Research Publication and Reviews*. Volume 2, Issue (8)(2021)ISSN 2582-7421, [www.ijrpr.com](http://www.ijrpr.com)

## Security IoT Device Against Emerging Security Threats: Challenges and Solution Techniques

**Madhumita Y. Sugandhi**

Student, Department of MCA, Vidya Bharati Mahavidyalaya, Amravati, India

**ABSTRACT** –The increasing prevalence of IoT device has brought about numerous security challenges due to their relatively simple internal architecture and low-powered hardware warranted by their small footprint requirement. The recent development in mobile computing results in widespread application of Internet of Things (IoT). IoT promises a world where smart and intelligent communication from most of the devices is possible through internet anywhere, anything with least possible human assistance. However, security and privacy are major concerns of IoT which could affect its sustainable development. Today's, the world is influenced by new emerging technologies. As a result we are surrounded by a number of smart devices. These smart devices make our life easy and convenient.

**KEYWORDS:** IoT Security; Cybersecurity; Security of threats; Sensors; Privacy; Internet.

### RESEARCH HIGHLIGHTS

- IoT devices pose significant security challenges due to their simple and low-footprints nature, which makes them incompatible with advanced technologies.
- The paper discusses common security threats, attacks associated with IoT devices and highlights the challenges associated with securing them.

### Introduction:

The Internet of Things (IoT) refers to a concept of connected object and devices of all types over the Internet wired or wireless. The popularity of IoT or the Internet of Things has increased rapidly, as these technologies are used for various purposes, including communication, transportation, education, and business development. IoT introduced the hyper connectivity concept, which means organizations and individuals can communicate with each other from remote location effortlessly.

IoT devices work by sending, receiving, and analyzing raw data from the real world, and are then used, either in part or wholly, to perform pre-programmed or user-defined action. Kevin Ashton invented the term, IoT in the years 1999 for promoting the Radio frequency Identification (RFID) concept, which includes embedded sensors and actuators. However the original idea was introduced in the 1960s. During that's period, the idea was called pervasive computing or embedded Internet. IoT has helped it to gain strong popularity in the summer of 2010. The Chinese government gave strategic priority on IoT by introducing a Five-Year plan. The mass explosion started in 2011 with the introduction of home automation, wearable devices and, smart energy meters. It is estimated that by the year 2030, there will be 25.44 billion active IoT devices worldwide, or in other word, three IoT devices for every single person on this planet.

The Internet of Things (IoT) has enhanced people's lifestyles by introducing automated services. However, this uncontrolled proliferation has led to heightened privacy and security challenges. Unconscious practices such as not changing passwords and neglecting device updates have escalated cybersecurity risks, allowing malicious applications to access sensitive data within IoT systems. These inadequate security measures significantly raise the likelihood of data breaches and other threats.

---

As the technologies underlying devices are becoming more sophisticated day by day, so is their use in our daily lives. This makes them an attractive target for cybercriminals as most IUT devices are inherently vulnerable due to their small or limited size capable of housing only low-power embedded microcontrollers, simple sensors, actuators, power supply units and other small electronic components. Memory and storage. This constraint of size coupled with low-power consumption requirements, simple or basic operating systems, and limited compute power often prevent the adoption of advanced or even modern cryptographic technologies, let alone complete security solutions. Also, much to the disappointment of cybersecurity experts and researchers, most IoT devices are still designed with only minimal built-in security, and this is especially true for cheap off-brand devices that come from both consumer and commercial markets.

Every day new technologies emerge, or changes are made to existing ones. Consider the latest advances in the 5G network, for example. 5G is expected to play an essential role in the IoT systems and applications. It is getting the researchers' attention and curiosity about the possible security and privacy risks, with its high frequency and bandwidth. Yet, the short wavelength imposes a change in the infrastructure, hence the need for more base stations to cover the same area covered by other wireless technology. This new structure imposes more threats, such as fake base stations. It is essential to understand the security risks and potential solutions. In this work, we aim to provide an overview of the IoT applications, benefits, and potential risks. Additionally, to build a framework to study and further develop best security practices by either implementing and analyzing current existing schemes or developing new ones. Based on the findings, we provide recommendations to avoid such risks and to remedy the possible security vulnerabilities. We built our model using Amazon Web Service (AWS) as proof of concept, which later translated to actual physical systems of sensors nodes mimicking general IoT structure. By making the system, we can deploy and study different security approaches by building real sceneries and benchmarks. The idea of IoT has made it easier for the world to connect devices in a way that is more accessible, reliable, available, scalable, private, and compatible.

### **Literature Review:**

The authors in said that there are challenges like jamming and fake attacks, as well as unauthorized access, which have affected the safety of user data. There are possible solutions to help people secure their IoT devices by implementing security measures. Mentions privacy threats that can impact IoT technologies and their networks. Managing security for IoT devices in businesses and organizations is not easy. Organizations need to use monitoring and scanning tools for IoT devices to detect and address privacy threats. Traffic interceptors and analyzers can identify and investigate cyber threats. Various studies and services, as mentioned in, have explored current trends in IoT security. They highlight challenges and attack vectors for different IoT devices and their guards. Simulation tools, modelers, and various platforms can help confirm security protocols for IoT.

According to, despite the benefits of the Internet of Things, there are challenges, especially in cybersecurity and privacy risks. These concerns affect both business and public organizations. Cybersecurity attacks on IoT technologies have exposed vulnerabilities due to the interconnected nature of IoT networks. Novel security solutions are needed to address accessibility from anonymous and untrusted internet sources. Emphasizing the standards and basic principles of the IoT Cyber Security Framework is crucial when implementing IoT security systems, as stated in one important measure is terminating contracts involving devices with different communication protocols.

## **Methodology & Structure:**

The methodology and structure of IoT (Internet of Things) security involve a comprehensive approach to safeguarding the interconnected devices and networks within the IoT ecosystem. Here's a breakdown of the key aspects:

### **Methodology:**

**Risk Assessment:** Identify potential threats and vulnerabilities in the IoT environment.

Evaluate the impact of security breaches on data integrity, confidentiality, and availability.

**Security by Design:** Integrate security measures during the design phase of IoT devices and systems. Implement secure coding practices to minimize vulnerabilities.

**Authentication and Authorization:** Establish robust mechanisms for device authentication and authorization. Employ strong, unique credentials for each IoT device.

**Data Encryption:** Encrypt data both in transit and at rest to protect sensitive information. Use industry-standard encryption algorithms.

**Network Security:** Implement firewalls and intrusion detection/prevention systems to secure IoT networks. Employ segmentation to isolate critical IoT components.

**Firmware and Software Updates:** Regularly update and patch firmware and software to address known vulnerabilities. Enable automated updates when possible.

**Monitoring and Incident Response:** Set up continuous monitoring for unusual activities. Develop an incident response plan to address security breaches promptly.

**Compliance with Standards:** Adhere to industry-specific and international security standards. Ensure compliance with data protection regulations.

### **Structure:**

**Device Layer:** Secure individual IoT devices with hardware-based security features. Implement device-level encryption and secure boot processes.

**Communication Layer:** Ensure secure communication protocols between devices and networks. Use protocols that support encryption and authentication.

**Cloud/Server Layer:** Secure cloud-based services and servers that store and process IoT data. Implement access controls and encryption for data storage.

**Gateway Layer:** Secure gateways that facilitate communication between IoT devices and the cloud. Apply security measures to prevent unauthorized access.

**User Interface Layer:** Implement secure user interfaces for monitoring and controlling IoT devices. Enforce strong authentication for user access.

**Lifecycle Management:** Establish secure processes for onboarding, provisioning, and decommissioning IoT devices. Monitor and manage security throughout the entire device lifecycle.

**Regulatory and Compliance Layer:** Ensure compliance with relevant regulations and standards. Implement controls to protect user privacy and data rights.

**Collaboration and Information Sharing:** Facilitate information sharing within the IoT security community to address emerging threats. Collaborate with stakeholders to enhance overall security posture. A robust IoT security strategy combines these elements to create a comprehensive framework that addresses the dynamic and evolving nature of security challenges in the IoT landscape. Regular updates and adaptation to emerging threats are critical to maintaining the effectiveness of IoT security measures.

This paper assumes that the reader is already well-versed in IoT devices and their underlying technologies. The qualitative and qualitative findings of this endeavor consisted of performing systematic literature review of &0 recently published worked in the field of cybersecurity, particularly IoT devices. The remainder of this paper.

### **Security Background:**

In simple terms, this text talks about computer security and how it has changed over time. In the 1980s, the main focus was on keeping information confidential, available, and intact. Confidentiality meant that only the sender and receiver should know about a message. Integrity meant that the content of a message should be the same for both the sender and receiver, and both parties could confirm this. Availability meant that the message should be readable by the sender and receiver at any time.

Later on, accountability was added as a security goal. This means that the sender and receiver should be able to prove the origin of a message, and the sender cannot send messages on behalf of someone else. The focus of computer security has shifted over the years to include confidentiality, integrity, accountability, and availability.

The text also introduces some terms. An adversary is someone who accesses a computer system's resources without permission. Malware is software that is designed to cause harm, like viruses. Risks in computer security occur when a threat (something that could breach security) and a vulnerability (a flaw in the system) come together. Even with perfect hardware and software, there would still be risks, so security measures are needed to minimize harm.

A threat is an event that could breach security and cause damage, but it needs the capability of execution or favorable circumstances. Vulnerability exists when there's a flaw in the system's design, implementation, operation, or management. An attack happens when all conditions are met, and it can be either active (altering system resources) or passive (gathering information). Attacks can come from inside or outside the system, with inside attackers having authorization to access the system but doing it in an unauthorized way. Outside attackers have no authorization to be inside the system.

### **Network Topologies:**

In simple terms, a computer network is like a group of computers connected to each other, allowing them to share information. The connection between computers is called a link, and it can be either through wires or wirelessly. The way these computers communicate with each other is governed by something called a network protocol. Networks make it possible for users to share resources and communicate with each other reliably and flexibly.

There are different types of telecommunication networks, and they can be categorized as physical or logical networks. Physical networks involve the actual physical devices like switches and routers connected to each other. In the past, the focus was mainly on securing these physical networks, but now the emphasis is on designing logical networks. Logical networks focus on how information flows between entities, regardless of the physical devices.

There are various network topologies, which describe how the computers are connected. One such topology is the point-to-point connection, where two computers are connected with a dedicated link. In this setup, both computers and the link between them need to work properly for the connection to function correctly.

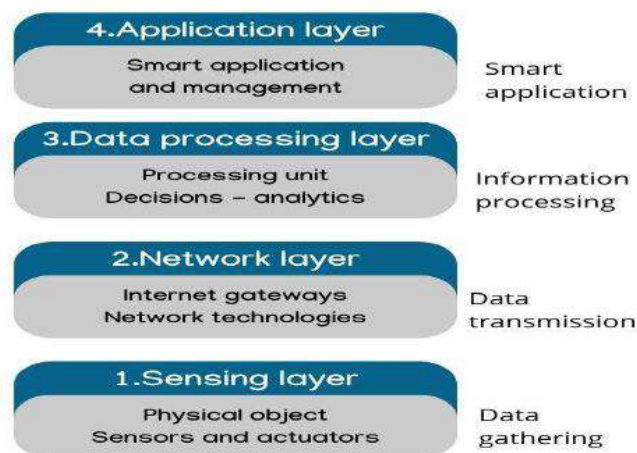


Fig 1. IoT Device Layers

### Sensing Layer:

The sensing layer is a crucial part of the IoT (Internet of Things) system that focuses on collecting data from the physical environment using various sensors. It serves as a middleman between the actual physical surroundings and the upper layers of the system, ensuring that the gathered data is precise, trustworthy, and relevant.

In the sensing layer, sensors play the key role of capturing information about the physical world. These sensors can include temperature sensors, humidity sensors, pressure sensors, light sensors, proximity sensors, and more. They are designed to measure specific physical quantities and convert them into electrical signals that the IoT system can process.

What makes the sensing layer special is its ability to collect both quantitative and qualitative data. For example, a temperature sensor can measure the exact temperature, while a camera sensor can capture images or videos. This variety of sensors allows the IoT system to gather a broad range of data types, providing a comprehensive understanding of the physical environment.



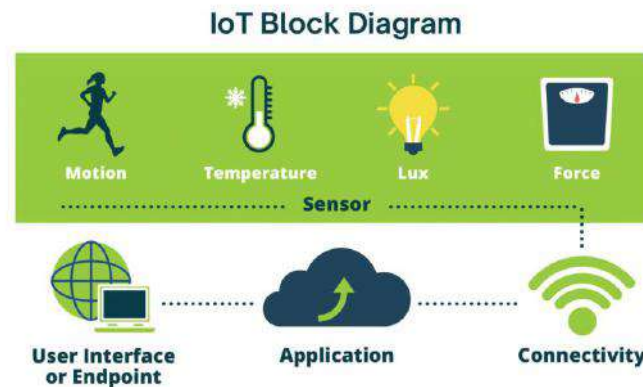


Fig 2. Sensing Layer

A significant challenge in the sensing layer is ensuring the accuracy and reliability of the collected data. Environmental conditions, sensor limitations, and signal interference can affect data quality. To address these issues, calibration, sensor maintenance, and data validation techniques are used to ensure the accuracy of the collected information.

Moreover, data fusion techniques are often applied in the sensing layer to combine data from multiple sensors, leading to more accurate and comprehensive insights. Integrating data from different sensors allows for a detailed understanding of the environment and facilitates complex decision-making processes.

The sensing layer acts as a link between the physical world and the digital realm, enabling the IoT system to collect real-time data and create a detailed picture of the physical environment. The accurate and reliable data gathered in this layer forms the basis for further analysis and decision-making in the higher layers of the IoT architecture.

### Networking Layer:

The networking layer of the IoT (Internet of Things) architecture is responsible for setting up connections and enabling smooth communication between devices and systems. It acts as a bridge that allows data transmission across various networks, enabling devices to communicate and share information.

In the networking layer, different communication protocols and technologies are used to establish connections between devices. These protocols include Wi-Fi, Bluetooth, ZigBee, cellular networks, Ethernet, and more. The choice of protocol depends on factors like range, data transfer rate, power consumption, and the specific needs of the IoT application.

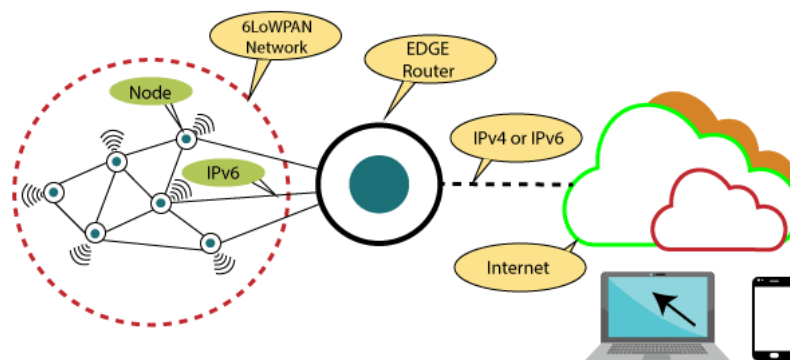


Fig 3. Network Layer

A major challenge in the networking layer is dealing with the diversity of devices and networks in the IoT ecosystem. Different devices may operate on different protocols or use different communication technologies. Therefore, protocols like MQTT (Message Queuing Telemetry Transport) and CoAP (Constrained Application Protocol) are employed to ensure interoperability and efficient communication between devices with varying characteristics.

Security is another crucial aspect of the networking layer. As data is transmitted across networks, it's essential to ensure the confidentiality, integrity, and availability of the information. Encryption algorithms, authentication mechanisms, and secure communication protocols are used to protect data from unauthorized access and maintain the integrity of the IoT system.

The networking layer plays a vital role in enabling scalability and flexibility within the IoT ecosystem. It facilitates the connection of devices, sensors, and actuators to form a network where data can be transmitted and shared seamlessly. This interconnected network of devices allows for real-time data exchange, remote control, and efficient collaboration.

In summary, the networking layer serves as the backbone of the IoT architecture, enabling the establishment of connections and smooth communication between devices. It plays a pivotal role in ensuring the efficient and secure transfer of data, which is crucial for the successful operation of the entire IoT system.

### Data Processing Layer:

In IoT sensor networks, wireless communication protocols are commonly used for exchanging information. These protocols operate in unlicensed frequency bands, providing flexibility and scalability for sensor deployments. However, using communication protocols for Wireless Sensor Networks (WSN) in unlicensed frequency bands can result in uncontrolled interference. This interference may lead to improper data transmission, resulting in sensor data with noise, missing values, outliers, and redundancy. This section explains various data analyses conducted to address issues in IoT sensor data, including denoising, missing data imputation, data outlier detection, and data aggregation.

The extensive sensor data generated in the IoT network requires data analysis, often with real-time decision-making. Sensor data possesses complex characteristics, including high velocities, large volumes, and dynamic values and types. Furthermore, sensor data can become contaminated with numerous obstacles until the necessary data analysis and real-time decision-making are achieved.

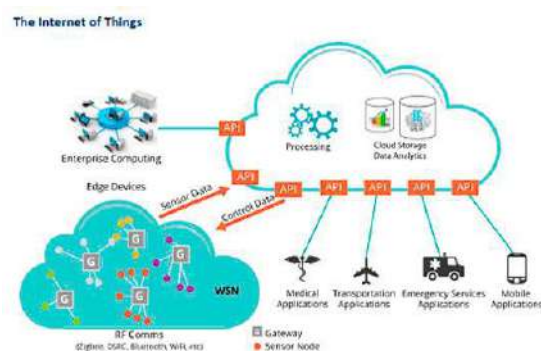


Fig 4. Data Processing Layer

Noise, in this context, refers to an uncorrelated signal component that introduces unwanted changes and modifications to the original vectors of the signal. Noise in the data necessitates unnecessary processing and resource utilization to handle unusable data. Wavelet transform methods prove effective in representing the signal and addressing the issue of signal estimation. Importantly, wavelet transformation preserves the original signal coefficients by eliminating noise within the signal. This is accomplished by thresholding the coefficients of

noise signals, making the perfect thresholding scheme essential. The wavelet transformation is a widely used method for analyzing and synthesizing the energy of continuous-time signals.

### Application Layer:

The application layer is the topmost layer of the IoT architecture and is responsible for providing specific functionalities and services tailored to the needs of end-users and organizations. It serves as the interface through which users interact with the IoT system and utilize the collected data for various applications and use cases.

These applications cover a wide range, including smart home systems, industrial automation, healthcare monitoring, agriculture management, and smart cities, among others. The applications in this layer offer value-added services, insights, and control to improve efficiency, productivity, and convenience.

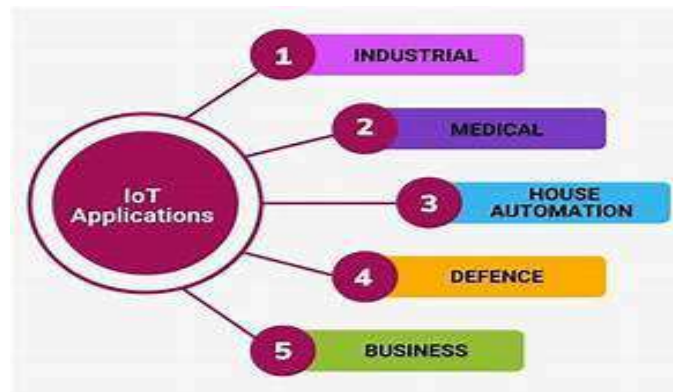


Fig 5. Application Layer

The application layer utilizes the data collected from the physical environment and processed by the lower layers to provide meaningful insights and actions. It involves activities such as data visualization, data analytics, and decision-making algorithms to extract valuable information and present it in a user-friendly and actionable manner.

User experience (UX) design is a key aspect of the application layer. IoT applications should be intuitive, easy to navigate, and offer a seamless user experience. This layer focuses on designing interfaces, dashboards, and controls that enable users to interact with the IoT system, monitor data, and initiate actions.

Additionally, the application layer often incorporates machine learning and artificial intelligence algorithms to enable predictive analytics and intelligent decision-making. The application layer also interacts with external systems and services, such as cloud platforms, third-party APIs, and other enterprise applications. This integration allows for extended functionality and interoperability, facilitating seamless data exchange and collaboration between different systems.

In summary, the application layer serves as the user-facing part of the IoT architecture, where specific use cases and functionalities are developed to meet the unique needs of end-users and organizations. It empowers users with valuable insights, control, and automation, enhancing various aspects of daily life and business operations.

### Result:

The outcomes or results of implementing IoT (Internet of Things) solutions can vary based on the specific goals, applications, and industries involved. Here are some common results or benefits associated with the implementation of IoT solutions:

**Increased Efficiency:** IoT enables the automation of various processes, leading to increased efficiency in operations. For example, in industrial settings, IoT can optimize production processes, reduce downtime, and improve overall efficiency.

**Cost Savings:** By leveraging IoT for monitoring and control, organizations can identify cost-saving opportunities. This may include predictive maintenance, energy management, and resource optimization, leading to reduced operational costs.

**Enhanced Productivity:** IoT solutions can contribute to improved productivity by providing real-time data and insights. This is particularly relevant in agriculture, manufacturing, and logistics, where data-driven decisions can enhance overall productivity.

**Improved Decision-Making:** The data collected by IoT devices can be analyzed to gain valuable insights, facilitating informed and data-driven decision-making. This is applicable across various sectors, such as healthcare, where IoT can aid in patient monitoring and diagnostics.

**Innovation and New Business Models:** IoT opens up opportunities for innovation and the creation of new business models. Companies can explore new revenue streams by offering IoT-enabled products or services, leading to business growth.

**Health and Wellness Benefits:** In healthcare, IoT devices can aid in remote patient monitoring, personalized treatment plans, and early detection of health issues, leading to improved health outcomes.

### **Conclusions:**

IT and control systems manufacturers are capitalizing on the emergence of new innovative hardware devices as the "Internet of Things" (IoT) continues to expand. With the increasing number of devices, there is a growing need for more automation in both consumer (e.g., home and car) and industrial environments. However, as automation in IoT control systems rises, so do software and hardware vulnerabilities. In the short term, data from IoT hardware sensors and devices will be managed by proxy network servers (e.g., a cellphone) due to the limited built-in security of current end devices and wearables. The security of these proxy devices becomes crucial when safeguarding sensor information. Eventually, the number of sensors per proxy will become large enough to make it inconvenient for users to manage them individually through separate apps. The future of IoT will require an exponentially larger volume of software to support the increasing number of devices. Despite this, the average number of software bugs per line of code remains unchanged, leading to an exponentially larger volume of exploitable bugs for adversaries. Until better standards for privacy protection, security guidelines, and data/cloud storage are established, the security of wearable and mobility devices will likely remain poor. The primary benefits of autonomous capabilities in the future IoT aim to extend and complement human performance. However, the danger of increased vulnerabilities is not being addressed at the same rate as innovation by security workers. The diversity of hardware and software in the future IoT presents both market competition and security challenges, as there is no single security architect overseeing the entire system. The lack of predefined governance, weak standards, and the dynamically defined nature of the IoT mission further emphasize the need for cooperation and collaboration between vendors for a secure future IoT, with no guarantee of success.

### **References:**

1. Khvoynitskaya, S. The History and Future of the Internet of Things. 2020. Available online: <https://www.itransition.com/>:<https://www.itransition.com/blog/iot-history> (accessed on 25 March 2020).
2. Conti, M.; Dehghantanha, A.; Franke, K.; Watson, S. Internet of Things security and forensics: Challenges and opportunities. *Future Gener. Comput. Syst.* 2018, 78, 544–546. [CrossRef]
3. Monther, A.A.; Tawalbeh, L. Security techniques for intelligent spam sensing and anomaly detection in online social platforms. *Int. J. Electr. Comput. Eng.* 2020, 10, 2088–8708.

4. Makhdoom, I.; Abolhasan, M.; Lipman, J.; Liu, R.P.; Ni, W. Anatomy of threats to the Internet of things. *IEEE Commun. Surv. Tutor.* 2018, 21, 1636–1675. [CrossRef]
5. Meng, Y.; Zhang, W.; Zhu, H.; Shen, X.S. Securing consumer IoT in the smart home: Architecture, challenges, and countermeasures. *IEEE Wirel. Commun.* 2018, 25, 53–59. [CrossRef]
6. Siby, S.; Maiti, R.R.; Tippenhauer, N.O. Iotscanner: Detecting privacy threats in IoT neighborhoods. In *Proceedings of the 3rd ACM International Workshop on IoT Privacy, Trust, and Security*, Abu Dhabi United Arab Emirates, 2 April 2017; pp. 23–30.
7. Hassan, W.H. Current research on Internet of Things (IoT) security: A survey. *Comput. Netw.* 2019, 148, 283–294.
8. Leloglu, E. A review of security concerns in Internet of Things. *J. Comput. Commun.* 2016, 5, 121–136. [CrossRef]
9. Liu, X.; Zhao, M.; Li, S.; Zhang, F.; Trappe, W. A security framework for the internet of things in the future internet architecture. *Future Internet* 2017, 9, 27. [CrossRef]
10. Ali, S.; Bosche, A.; Ford, F. *Cybersecurity Is the Key to Unlocking Demand in the Internet of Things*; Bain and Company: Boston, MA, USA, 2018. *Appl. Sci.* 2020, 10, 4102 16 of 17
11. Sadeghi, A.-R.; Wachsmann, C.; Waidner, M. Security and privacy challenges in industrial internet of things. In *Proceedings of the 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, San Francisco, CA, USA, 8–12 June 2015; pp. 1–6.
12. Izzat, A.; Chuck, E.; Lo'ai, T. *The NICE Cyber Security Framework, Cyber Security Management*; Springer: Basel, Switzerland, 2020; ISBN 978-3-030-41987-5.
13. Tawalbeh, L.A.; Tawalbeh, H. Lightweight crypto and security. In *Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications*; Wiley: West Sussex, UK, 2017; pp. 243–261.
14. Alaba, F.A.; Othman, M.; Hashem, I.A.T.; Alotaibi, F. Internet of Things security: A survey. *J. Netw. Comput. Appl.* 2017, 88, 10–28. [CrossRef]
15. Conti, M.; Dragoni, N.; Lesyk, V. A survey of man in the middle attacks. *IEEE Commun. Surv. Tutor.* 2016, 18, 2027–2051. Available online: <https://ieeexplore.ieee.org/abstract/document/7442758> (accessed on 10 April 2020). [CrossRef]
16. Khan, M.A.; Salah, K. IoT security: Review, blockchain solutions, and open challenges. *Future Gener. Comput. Syst.* 2018, 82, 395–411. [CrossRef]

## Using Machine Learning to Enhance Security Measures at the Network Layer of IoT

**Ms. Varkha K. Jewani**  
(**Ms. Pragati V. Thawani**)

M.Sc. (IT), M.Phil. (IT)  
Assistant Professor  
K.C College, Churchgate  
vkjewani@gmail.com

**Dr. Prafulla E. Ajmire**

M.Sc., PGDCS, M.S. (Soft.Syst.), M.Phil. (CS), Ph.D.  
Head & Associate Professor in Computer Science & Application  
G S Science, Arts & Commerce College, Khamgaon, Maharashtra  
Sant.Gadge Baba Amravati University, Maharashtra  
peajmire@gmail.com, peajmire@rediffmail.com

**Ms. Geeta N. Brijwani**

M.Sc. (CS), M.Phil. (CS)  
Assistant Professor  
K.C College, Churchgate  
geetabrijwani@gmail.com

**Abstract:** The Internet of Things, or IoT, has expanded quickly as a means of increasing ease and efficiency by connecting a wide range of objects into networks. However, there are serious security risks associated with this pervasive networking. To reduce potential threats and vulnerabilities, this abstract examines the critical necessity for improving security measures at the Internet of Things' network layer. The network layer plays a crucial role on the Internet of Things' communication infrastructure by enabling smooth integration and facilitating data transmission between devices. It is, nevertheless, vulnerable to several security risks, including as denial-of-service attacks, illegal access, and data leaks. This study suggests a thorough strategy for enhancing network layer security to protect the availability, integrity, and confidentiality of Internet of Things systems. This paper proposes a multifaceted and comprehensive solution to improve security at the Internet of Things' network layer. Organizations and IoT stakeholders can create a strong security framework that guarantees the IoT ecosystem's continuous growth and success by tackling vulnerabilities at their core. The suggested actions provide a safe and dependable Internet of Things ecosystem by protecting sensitive data and helping to increase user trust.

**Keywords:** IoT, Network Layer, DDos Attack, Machine Learning.

### I. INTRODUCTION

The rapid growth of the Internet of Things (IoT) has revolutionized the ways in which gadgets interact, cooperate, and support different facets of everyday life and business processes. The Internet of Things (IoT) has brought about previously unheard-of levels of efficiency and convenience, from smart homes to networked industrial systems. But increased connectivity has also brought forth a host of new security issues, with the network layer acting as a vital front in the fight against possible attackers.

The network layer is the backbone of the Internet of Things, allowing data interchange and smooth communication amongst a wide range of devices. Although there is unmatched potential brought about by this interconnection, there are also weaknesses introduced that could be used by bad actors. Network security lapses put consumers' privacy and the integrity of IoT ecosystems at danger since they can result in tampering, data manipulation, and unapproved access. The circumstances for the investigation of the essential need to improve security measures at the Internet of Things' network layer is established by this introduction. The network layer requires more attention to strengthen its defenses against emerging cyber threats because it is the key that connects devices and allows them to collaborate. This conversation seeks to offer practical solutions that businesses, developers, and regulators may implement to guarantee the resilience and integrity of IoT networks by exploring creative and all-encompassing security tactics. Enhancing network security involves addressing a variety of issues, from enforcing strong authentication procedures and safeguarding communication

protocols to utilizing cutting-edge technology like blockchain. This article seeks to establish a secure foundation for the long-term growth and success of IoT ecosystems by promoting a proactive and comprehensive strategy that goes beyond traditional security procedures. As we continue our investigation, it becomes clear that strengthening the network layer is not only a matter of technical necessity but also a critical first step toward establishing longevity, dependability, and trust in the rapidly changing IoT world.

## II. LITERATURE REVIEW

- **Title: "Security Challenges in the Internet of Things: A Comprehensive Survey"**  
Authors: Antonios Gouglidis, Helge Janicke, Ioannis Mavridis This survey provides a comprehensive overview of security challenges in IoT. It emphasizes the significance of securing the network layer and discusses various threats and vulnerabilities. The authors stress the need for encryption, secure communication protocols, and intrusion detection mechanisms.
- **Title: "A Survey on Security in Internet of Things: From Device to Fog and Cloud"**  
Authors: Pradeep Kumar Tiwari, Ching-Hsien Hsu, Anand Paul This survey explores security issues across the entire IoT ecosystem, with a focus on the network layer. It discusses the importance of secure communication, authentication, and access control mechanisms. The authors highlight the role of blockchain and machine learning in enhancing security at the network layer.
- **Title: "Securing the Internet of Things: A Standardization Perspective"** Authors: O. Vermesan, P. Friess Focusing on standardization efforts, this work reviews security measures for IoT, including those at the network layer. It discusses the importance of standardized security protocols, such as CoAP and DTLS, and emphasizes the need for interoperability to ensure consistent security across diverse IoT devices and networks.
- **Title: "Blockchain-Based Security Framework for IoT"** Authors: Ali Dorri, Salil S. Kanhere, Raja Jurdak, Praveen Gauravaram This paper explores the role of blockchain in enhancing security for IoT, particularly at the network layer. It discusses how blockchain's decentralized and tamper-resistant nature can be leveraged to secure communication, prevent unauthorized access, and establish trust among IoT devices.
- **Title: "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications"** Authors: Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu Palaniswami This survey provides a comprehensive overview of IoT, covering various aspects including security. It discusses security challenges at the network layer and proposes solutions such as network segmentation, secure communication, and the importance of regular updates to mitigate vulnerabilities.
- **Title: "Security in the Internet of Things: A Review"** Authors: Francesco B. Gaetani This review article focuses on security issues in IoT, highlighting challenges and proposing solutions. It emphasizes the need for secure communication protocols and access control mechanisms to protect the network layer from unauthorized access. The paper also discusses the role of encryption and key management in securing IoT networks.
- **Title: "A Survey on Internet of Things Architectures"** Authors: N. Baccour, M. K. Marina, M. Khan, M. Z. Shafiq, A. Alomainy This survey delves into various IoT architectures and their security implications. It emphasizes the role of the network layer in enabling communication and data exchange and discusses security challenges such as

eavesdropping and man-in-the-middle attacks. The authors propose secure communication protocols and network segmentation as crucial measures.

The literature on the enhancement of security measures at the network layer of IoT underscores the importance of adopting a multi-faceted approach. Secure communication protocols, authentication mechanisms, blockchain, and standardization efforts emerge as key components in fortifying the network layer against evolving cyber threats in the dynamic landscape of IoT.

### III.METHODOLOGY

- 1) **Risk Assessment and Threat Modeling:** To find potential threats and vulnerabilities unique to the network layer of the Internet of Things environment, do a thorough risk assessment.
- 2) **Security Protocol Analysis:** Examine current IoT network layer communication protocols for flaws and vulnerabilities. To secure data in transit, choose and implement reliable security protocols like Datagram Transport Layer Security (DTLS), Transport Layer Security (TLS), or other appropriate options.
- 3) **Authentication Mechanisms:** To guarantee that only authorized devices can access the network, implement robust authentication procedures, such as mutual authentication. To improve the overall security posture, investigate the usage of certificate-based authentication and secure key exchange methods.
- 4) **Authorization Policies:** Create and implement granular authorization policies to regulate the behaviors that are allowed for every Internet of Things device connected to the network. To guarantee that devices have access to only the resources and functions required for their intended purposes, implement role-based access control, or RBAC.
- 5) **Intrusion Detection and Prevention System:** Implement a network based IDPS to track and examine network activity in real time. Install anomaly detection tools to find odd behavior patterns and possible security breaches. Put in place automated reaction systems to quickly neutralize dangers that are identified.
- 6) **Blockchain Integration:** Examine whether incorporating blockchain technology into the IoT network layer can improve transaction security and transparency. Discover how blockchain technology may be used for tamper-resistant transaction verification, safe and decentralized identity management, and secure data provenance.
- 7) **Network Segmentation:** Use network segmentation to divide up the various IoT infrastructure segments. Establish and implement stringent access rules between network segments to minimize the effect of possible security incidents and stop lateral network movement.
- 8) **Firmware and Software Updates:** Provide a reliable and safe system for updating IoT devices' firmware and software on a regular basis. To guarantee that devices are running the most recent security patches and mitigations against known vulnerabilities, use automated updating procedures.
- 9) **Security Education and Training:** Assist IoT network administrators, developers, and end users with continual security education and training. To promote responsible behavior and improve the IoT ecosystem's overall cybersecurity resilience, cultivate a culture that is security aware.
- 10) **Regular Security Audits and Assessments:** To find and fix new security vulnerabilities, conduct routine security audits and evaluations of the IoT network layer. To guarantee a comprehensive analysis of the network's security posture, use outside security professionals to perform penetration testing and vulnerability assessments.

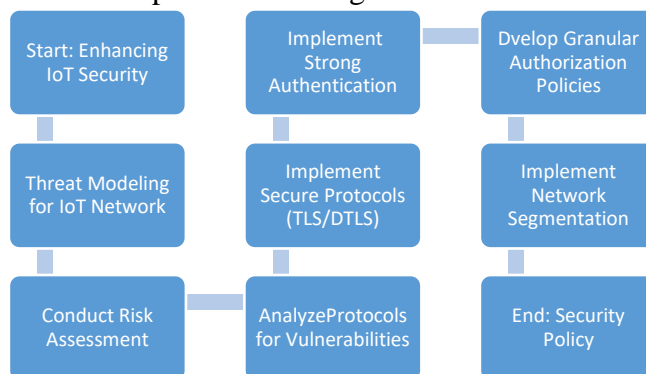


11) **Collaboration with Standards Bodies:** Keep up with changes in IoT-related security standards and best practices. Work together with industry standards organizations to guarantee adherence to new security guidelines and support the creation of safe Internet of Things procedures.

By systematically applying this methodology, organizations can strengthen the security measures at the network layer of their IoT ecosystems, thereby reducing the risk of unauthorized access, data breaches, and other security threats.

#### IV. FLOW DIAGRAM

Creating a flow diagram for enhancing security measures at the network layer of IoT involves outlining the key steps and components involved in securing the network layer. Below is a simplified flow diagram that one can use as a starting point.



**Fig: 1- Framework of Security measures at Network Layer**

This flow diagram outlines the sequential steps involved in enhancing security at the network layer of IoT. Each box represents a specific action or process, and arrows indicate the flow of the overall enhancement process. This visualization serves as a guide for organizations seeking to fortify their IoT networks against potential security threats and vulnerabilities.

#### V. MACHINE LEARNING ALGORITHMS

Machine learning (ML) algorithms play a crucial role in enhancing security measures at the network layer of IoT by enabling intelligent threat detection, anomaly identification, and real-time response. Below are some machine learning algorithms that can be employed for different aspects of network layer security in IoT:

##### 1. Intrusion Detection:

###### Algorithm: Random Forest

**Description:** Random Forest is an ensemble learning algorithm that combines the outputs of multiple decision trees. It is effective for detecting anomalies and identifying patterns indicative of network intrusions.

###### Algorithm: Support Vector Machine

**Description:** SVMs are powerful for binary classification tasks and can be used to detect abnormal network behavior. SVMs are effective when dealing with high-dimensional data, making them suitable for network traffic analysis.

##### 2. Anomaly Detection:

###### Algorithm: Isolation Forest

**Description:** The Isolation Forest algorithm is designed for anomaly detection. It isolates anomalies by constructing random decision trees and measuring the number of splits required to isolate a data point. It is efficient and effective for detecting unusual patterns in network traffic.

###### Algorithm: Autoencoders (Neural Networks)

**Description:** Autoencoders, a type of neural network, can learn the normal patterns of

network traffic and identify anomalies by detecting deviations from the learned representations. They are particularly useful for capturing complex, non-linear relationships in data.

### 3. Behavioral Analysis:

#### **Algorithm: Hidden Markov Models (HMM)**

##### **Description:**

HMMs are suitable for modeling the sequential nature of network activities. They can capture the transitions between different states of network behavior and detect deviations from expected patterns.

#### **Algorithm: Long Short-Term Memory (LSTM)**

##### **Networks**

**Description:** LSTM networks, a type of recurrent neural network, excel at capturing temporal dependencies. They are effective for analyzing sequences of network events and identifying abnormal patterns that may indicate security threats.

### 4. Threat Intelligence Integration:

#### **Algorithm: Naive Bayes Classifier**

**Description:** Naive Bayes classifiers are simple and efficient for integrating threat intelligence data into security analysis. They can categorize network activities based on known threat indicators and aid in real-time decision-making.

### 5. Network Traffic Classification:

#### **Algorithm: K-means Clustering**

**Description:** K-Means clustering can be used for classifying network traffic into different categories. By grouping similar traffic patterns, it becomes easier to identify and prioritize potential security issues.

### 6. Predictive Analysis for Vulnerability Management:

#### **Algorithm: Decision Tree**

**Description:** Decision trees are useful for predicting vulnerabilities in the network layer based on historical data. By analyzing features related to past security incidents, decision trees can provide insights into potential future risks.

### 7. Multi-Layered Defense:

#### **Algorithm: Ensemble Methods (e.g., Stacking)**

**Description:** Combine multiple machine learning models, each specializing in different aspects of security, into an ensemble. Stacking, for example, integrates the predictions of diverse models, enhancing overall accuracy and robustness.

These machine learning algorithms can be tailored to specific use cases within the enhancement of security measures at the network layer of IoT. It's essential to continuously train and update these models to adapt to evolving threats and maintain effectiveness over time. Additionally, the choice of algorithm may depend on the specific characteristics of the IoT network and the nature of the security challenges it faces.

## VI CONCLUSION

In conclusion, a viable approach to mitigating the dynamic and ever-evolving nature of cyber threats is to use machine learning (ML) to improve security measures at the network layer of the Internet of Things (IoT). By incorporating machine learning algorithms into security frameworks, intelligence and flexibility are added, resulting in enhanced threat detection, anomaly detection, and network protection. Though there is promise in integrating ML with IoT network security, there are a few things to keep in mind. ML models need to be updated and trained continuously in order to respond to changing threats. Additionally, it might be difficult to comprehend the reasoning behind some security alerts due to the interpretability of machine

learning choices, particularly in sophisticated neural networks. In conclusion, applying machine learning (ML) to improve IoT network security measures is an essential first step toward developing flexible, astute, and potent defenses against the always evolving array of cyber threats. As technology develops, continued study and research into machine learning applications for IoT security will further hone and enhance these systems' effectiveness.

## ACKNOWLEDGMENT

The author is very much thankful to Dr. P.E Ajmire sir for all his guidance and IEEE for giving this opportunity to publish research paper for this particular conference and all the members of conference team.

## References

- [1] N. Moustafa, B. Turnbull, K.-K.R. Choo, Towards automation of vulnerability and exploitation identification in IIoT networks, in: 2018 IEEE International Conference on Industrial Internet. DOI:10.1109/ICII.2018.00023 Corpus ID: 53948805
- [2] N. Moustafa, B. Turnbull, K.-K.R. Choo, an ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of IoT, IEEE Internet Things J. (2018). DOI:10.1109/ICII.2018.00023
- [3] S. Salnyk, A. Storchak, A. Mykytyuk, Information Technology and Security 7 (2019), Iss. 1, pp. 25-34.
- [4] M. Tayyab, B. Belaton, and M. Anbar, "ICMPv6 based DoS and DDoS attacks detection using machine learning techniques, open challenges, and blockchain applicability: a review," IEEE Access 8 (2020), pp. 170529–170547. DOI: 10.1109/ACCESS.2020.3022963.
- [5] Ahmed, Sheikh, A Study of ML Algorithms for DDoS Detection, International Journal for Research in Applied Science and Engineering Technology (2021).
- [6] Y. Alshboul, and K. Streff, "Analyzing Information Security Model for Small-Medium Sized Businesses", in Proc. 21st Americas Conference on Information Systems, Puerto Rico, 2015.
- [7] M. A. Simplício, M. V. M. Silva, R. C. A. Alves, and T. K. C. Shibata, "Lightweight and escrowless authenticated key agreement for the internet of things", Comput. Commun. 98 (2017), pp.43–51.
- [8] J. Czyz, M. J. Luckie, M. Allman, and M. Bailey, "Don't forget to lock the back door! A characterization of IPv6 network security policy", in NDSS, 2016.
- [9] K. Angrishi, "Turning internet of things (IoT) into internet of vulnerabilities (IOV): IoT botnets", CoRR, vol. abs/1702.03681, 2017. arXiv: 1702.03681. URL: <http://arxiv.org/abs/1702.03681.236>
- [10] D. Parwani, A. Dutta, P. Kumar Shukla, et al, Various techniques of DDoS attacks detection and prevention at cloud: a survey, Orient. J. Comp. Sci. and Technol. 8(2015), no. 2, URL: <http://www.computerscijournal.org/?p=1983>
- [11] Jadel Alsemiri & Khalid Alsubhi, Internet of Things Cyber Attacks Detection using ML, International Journal of Advanced Comp. Science and Applications (2019).
- [12] Nour Moustafa, Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic, PhD thesis, University of New South Wales, Canberra, Australia.
- [13] K. Nickolaos, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the IoT for network forensic analytics: Bot- IoT dataset", Future Generation Comp. Systems 100 (2019).
- [14] Mehryar Mohri, Afshin Rostamizadeh, & Ameet Talwalkar, Foundations of Machine Learning, 2nd.ed., the MIT Press, 2018.

## The Role of chatGPT and other Artificial intelligent in the field of Renewable Energy and sustainable Energy

**Mr.Anup Satishrao Bele**

[anupsbele@gmail.com](mailto:anupsbele@gmail.com) 9960736402

Institute Of Management Studies Mahavidyalaya, Warud

### **Abstract:**

The integration of artificial intelligence (AI) technologies like ChatGPT in the renewable energy sector opens up more opportunities to improve efficiency, reduce operations and advance sustainability goals. Here, we explore potential applications of ChatGPT in renewable energy, focusing on areas as diverse as smart grids, energy management systems, predictive maintenance, and customer interaction.

ChatGPT has the potential to change the way the energy industry operates and will help improve energy management in the future, support the development of renewable energy, support the development of new energy technologies and also help the energy market.

The energy industry is constantly looking for ways to improve efficiency and reduce environmental impact. Using ChatGPT's capabilities, energy companies can gain insights and make more informed decisions that increase efficiency and sustainability. This article takes a closer look at how ChatGPT is used in the energy industry and how it can potentially benefit the industry. From automated data analysis to improving power plant efficiency to supporting the development of renewable energy sources, ChatGPT has proven to be an invaluable tool for achieving a more sustainable future. Let's take a closer look at how this cutting-edge technology is shaping the future of the energy industry.

ChatGPT is a powerful tool that can be used to process and analyze large amounts of data and generate human-like text. This makes it a valuable asset in the energy industry, where efficiency and sustainability are key concerns.

**Keywords:** *ChatGPT, artificial intelligence, renewable energy, energy management, predictive maintenance.*

### **1. Introduction**

ChatGPT became available to the public the internet has been buzzing with excitement. The chatbot - and OpenAI, the company behind its development - is extremely impressive and will disrupt many industries in ways that we couldn't imagine before. I was interested what impact it could have on the energy industry and how could it support the energy transition

ChatGPT is a state-of-the-art language model that can be fine-tuned to do a variety of tasks, such as answering questions, writing text, and even creating code. One of the key capabilities of ChatGPT is its ability to understand and process large amounts of data, which makes it an ideal tool for automating data analysis.

Automating data analysis in power plants ChatGPT can be used to improve the efficiency of power plants by automating data analysis. Power plants generate a large amount of data, including information about energy production, equipment performance, and environmental conditions. Analyzing this data can provide valuable insights into how to improve the efficiency of the power plant and reduce its environmental impact.

Traditionally, data analysis in power plants is done manually, which can be time-consuming and prone to errors. However, by using ChatGPT, energy companies can automate the data analysis process, making it faster and more accurate. ChatGPT can be trained to understand the

data generated by power plants and identify patterns and trends that can indicate inefficiencies or potential problems. This can help energy companies make more informed decisions about how to improve the performance of the power plant, such as by identifying equipment that needs maintenance or by optimizing energy production.

Additionally, ChatGPT can also assist in monitoring and reporting on the environmental impact of power plants. By analyzing data on emissions and other environmental factors, ChatGPT can help energy companies identify ways to reduce the environmental impact of their operations.

ChatGPT can be used to support the development and implementation of renewable energy sources. Renewable energy sources, such as solar, wind, and geothermal, are becoming increasingly important as the world looks for ways to reduce its dependence on fossil fuels and combat climate change. However, developing and implementing these energy sources can be a complex and challenging task.

ChatGPT can assist in the development of renewable energy sources in a number of ways. For example, it can be used to analyze data on weather patterns and wind speeds to identify the best locations for wind turbines.

It can also be used to generate reports and simulations to help engineers design and optimize the performance of renewable energy systems. Additionally, ChatGPT can also be used to assist in the development of new renewable energy technologies, such as by analyzing large amounts of research data to identify new possibilities or by creating machine learning models to predict performance of new technologies.

### 1.1 Top 10 Green Energy Innovation Trends

1. Advanced Photovoltaic's <b>19%</b>	2. AI and Big Data <b>19%</b>	3. Distributed Energy Storage Systems <b>16%</b>	4. Hydropower <b>16%</b>	5. Wind Energy <b>11%</b>
6. Bioenergy <b>5%</b>	7. Grid Integration <b>4%</b>	8. Green Hydrogen <b>4%</b>	9. Advanced Robotics <b>3%</b>	10. Blockchain <b>3%</b>

Table 1.0

## 2 .MATERIALS AND METHODS

The integration of ChatGPT in the field of renewable energies provides new approaches to several important aspects, especially in research, analysis and communication. Below are materials and methods that define ChatGPT's role in promoting renewable energy.

### 2.1 Materials:

ChatGPT Model: The core content is the ChatGPT model, which is trained on large datasets including renewable energy literature, technical documents, policies and discussions.

Renewable Energy Data: Datasets containing information about renewable energy technologies.

Research Papers and Publications: Access to a wide range of renewable energy related research papers, journal articles and publications enables ChatGPT to stay updated with the latest advances, methods and findings in the field.

### 2.2 Methods:

Natural Language Understanding (NLU):

ChatGPT uses advanced natural language understanding techniques to understand user queries, extracting key information, context and intent related to renewable energy topics.

Information retrieval and synthesis:

ChatGPT derives relevant information from renewable energy datasets, research papers and publications to effectively address user queries.

**Problem solving and decision support:**

ChatGPT helps researchers, policy makers and industry professionals in problem solving and decision making processes related to renewable energy.

**Interactive Dialogue and Learning:**

Through interactive dialogue, ChatGPT engages users in educational conversations about renewable energy technologies, policies and sustainability principles.

**Innovation and Idea Generation:**

ChatGPT serves as a catalyst for innovation in renewable energy by generating novel ideas, exploring alternative approaches and fostering creative thinking among researchers, engineers and entrepreneurs.

### **3. Conclusions**

Conclusions and overall impact on efficiency and sustainability in the energy industry.

We discussed how ChatGPT can be used in various ways to improve efficiency and sustainability in the energy industry, such as automated data analysis in power plants, helping the development of renewable energy sources, improving energy efficiency in buildings and homes, streamlining. Communication and decision making in energy companies and identifying potential future applications of ChatGPT in the energy industry.

The integration of ChatGPT into renewable energy applications holds immense promise to revolutionize how we generate, distribute and use clean energy resources. By leveraging its cognitive capabilities, natural language understanding and data analysis skills, ChatGPT can catalyze innovation, drive efficiencies and accelerate the global transition to a more sustainable energy future.

### **4.Reference**

- 1) "Artificial Intelligence for Sustainable Energy Production and Consumption" edited by Valentina Emilia Balas, Lakhmi C. Jain, and Marius Mircea Balas
- 2) "Renewable Energy: Power for a Sustainable Future" by Godfrey Boyle, Bob Everett, Janet Ramage, and Stephen Peake

## **An In-Depth Exploration of AI in the Digital Age- A Focus on Scams and Frauds in the Introduction Phase**

**Ms. Rhutika V. Jawarkar**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
[rhutikajawarkar25@gmail.com](mailto:rhutikajawarkar25@gmail.com)

**Dr. Priyanka C. Tikekar**

Department of Computer Science, Bharatiya Mahavidyalaya, Amravati, MS, India  
[priyanka.tikekar058@gmail.com](mailto:priyanka.tikekar058@gmail.com)

### **Abstract-**

This study explores sneaky tricks used in artificial intelligence (AI) scams, revealing how bad actors cleverly misuse AI for scams. As AI becomes more widespread, the dark side of using smart technology for dishonest activities is also growing, putting people, businesses, and online security in danger. The research looks into secret networks and places where these AI scams are planned and carried out. It emphasizes the importance of better online security and teamwork between countries to tackle these global threats. The paper also talks about the ethical problems linked to using AI for scams, making people more aware of the potential harm caused by misusing powerful technologies. paper provides helpful information about the changing world of AI-driven scams. It lays the groundwork for creating strong defenses and rules to protect people and organizations from the increasing dangers caused by those who misuse artificial intelligence.

*Keywords-* Artificial intelligence, AI scams, frauds

### **I. INTRODUCTION**

Artificial Intelligence is a technology that allows machines to think and act like humans. It helps them learn, solve problems, and do various tasks. AI has brought many advantages in industries like technology, banking, marketing, and entertainment. For example, it's used in driving cars, setting fitness goals on smartwatches, suggesting songs and shows on streaming services, and planning the best travel routes with map apps [1]. However, there's a downside. Scammers are using AI in a tricky way. They gather personal information from social media and other online sources. Then, they use this info with AI to customize scam texts and emails. This makes the scams look real and harder to spot as fake. Scammers are playing a big game, and people can easily fall into their traps [2].

### **II. ARTIFICIAL INTELLIGENCE SCAMS**

Artificial Intelligence (AI) refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, language understanding, and decision-making. The goal of AI is to create machines that can mimic and replicate human cognitive functions, ultimately achieving a level of intelligence and problem-solving capability comparable to or beyond that of humans. Artificial Intelligence, makes it simpler for scammers to create scams. They use different techniques to trick people. Scammers make scams that involve personal information,

---

money, and banking frauds. They play tricks to trap unsuspecting individuals. Some AI scams are discussed below [3].

*A. Generating text and image content*

Using AI, scammers can make personalized emails, instant messages, and images to trick people. They use these to lure victims into scams like phishing or fraudulent advertisements. Traditional signs of scams, like bad spelling and grammar, can be fixed using AI, making scams harder to spot [4]. AI can also be used to create fake images, like pictures of damaged cars or property, to support fake insurance claims. While some AI tools have safety features, our research found that scammers can bypass them. People we talked to in our research think this threat is already a big problem and will keep growing [5].

*B. AI enabled chatbots*

Some organizations talked to have found proof that clever chatbots, possibly using AI, are talking to potential scam victims. They try to manipulate people into paying money. In some cases, scammers purposely ask confusing questions in chats to gather information. This leads to strange responses that seem like they're from a computer.

Chatbots can help scammers reach more victims without needing many people. Just one chatbot could trick as many people as a whole call Center could before. Though there's not a lot of proof that this is happening a lot right now, the people we talked to in our research think this problem will get worse. Chatbots will make scams more common, and it will be harder to tell if it's a computer or a person talking [6].

*C. Deep fakes video*

Fake videos, known as deep fakes, are already tricking people. Some use them as "click bait" to make people visit harmful websites that steal credit card details for fraud. There have been cases where scammers use the image and identity of someone trustworthy to get people to interact with harmful content. A bank using "selfie video ID" told us about basic attempts where people showed a deep fake video to their webcam, trying to trick the system. Deep fake videos will become more common. People we talked to are watching closely to see how technology and changes in behavior can help individuals tell what content is trustworthy and what is not [7].

*D. Voice cloning-scams and voice ID*

Deep fake technology can copy voices very accurately. Right now, making a really good voice clone might need a couple of hours of audio, but this might not be necessary in the future. A tech company we talked to thinks that soon, a voice clone that sounds just like a real person could be made with only a few seconds of audio. Scammers are already using voice cloning in scams. For instance, they might use a cloned voice to leave a voicemail, or even have a live conversation with someone pretending to be someone else. This could be used for fraud easily. Imagine someone getting a call from a person sounding like their boss, asking them to send money to a different account [8].

### **III. FRAUD DETECTION**

Fraud detection refers to the process of identifying and preventing fraudulent activities within various systems, transactions, or processes. Detecting fraud is crucial in safeguarding individuals, businesses, and organizations from financial losses, reputational damage, and legal consequences [9].



- **Assessment of Fraud Detection Methods:** Traditionally, fraud detection relied on rule-based systems with predefined rules to identify suspicious transactions. However, these systems often faced challenges with high false positive rates and struggled to adapt to new and emerging fraud patterns.
- **Integration of Artificial Intelligence in Fraud Detection:** The application of Artificial Intelligence (AI) techniques, including Neural Networks and decision trees, has demonstrated significant improvements in fraud detection accuracy. For example, Zheng et al. (2020) introduced a deep learning-based fraud detection framework that outperformed traditional methods.
- **Role of Machine Learning in Fraud Detection:** Machine learning algorithms have become widely utilized in fraud detection due to their capability to identify intricate patterns and anomalies in large datasets. In a study by Johnson et al. (2019), a fraud detection system was developed using a random forest algorithm.
- **Real-Time Fraud Detection:** Swift identification of fraud in real-time is crucial for preventing financial losses and mitigating the impact of fraudulent activities. Deep learning models, such as recurrent Neural Networks, have been employed for real-time fraud detection owing to their ability to process sequential data.
- **Deep Learning Architectures for Fraud Detection:** Advanced deep learning architectures, such as Neural Networks and Convolutional Neural Networks, have shown promise in enhancing fraud detection capabilities. These sophisticated models offer improved accuracy and effectiveness in identifying fraudulent activities [10].

#### **IV. OTHER TYPES OF FRAUDS**

##### *A. Credit card fraud*

Credit card fraud is a widespread form of cybercrime where fraudsters use various methods to obtain users' personal data. They may hack into this information through phone calls, Wi-Fi hotspots, or emails. Importantly, these fraudsters can steal the cardholder's details without physically having the card in their possession, making online means their primary tool for carrying out such activities [11].

##### *B. Mobile fraud*

Mobile fraud refers to deceptive and illicit activities conducted through mobile devices, typically involving the unauthorized access, manipulation, or exploitation of personal or financial information. This type of fraud exploits vulnerabilities in mobile technologies and takes advantage of users' trust in their smartphones or other mobile devices.

##### *C. Internal fraud*

Internal fraud refers to dishonest and deceptive activities carried out by individuals within an organization. Unlike external threats, internal fraud involves employees, contractors, or other individuals with direct access to the organization's systems, assets, or sensitive information. Internal fraud can take various forms, and it often involves the misuse or theft of resources for personal gain.

##### *D. Money laundering*

Money laundering is the process of making illegally obtained gains, often from criminal activities, appear legitimate by passing them through a complex sequence of banking transfers

or commercial transactions. The primary goal of money laundering is to obscure the true origin of the funds, making them appear to be derived from legal sources.

### *E. Identity and social fraud*

Identity and social fraud involve deceptive practices where individuals or entities misrepresent their identity or manipulate social relationships for personal gain or to engage in fraudulent activities. These fraudulent activities can have various forms and implications [12][13].

## V. CONCLUSION

In conclusion, the exploration of AI scams and frauds underscores the complex and evolving challenges presented by the intersection of technology and deceptive practices. The rapid advancement of artificial intelligence has not only ushered in innovative solutions but has also provided fertile ground for malicious actors to orchestrate sophisticated scams. AI scams, ranging from personalized fraudulent content to deepfake videos and voice cloning, highlight the adaptability of scammers in exploiting cutting-edge technologies. These scams capitalize on the vast pool of personal information available online, making them not only more convincing but also harder to detect. As technology continues to progress, the risks associated with AI scams demand heightened awareness, robust cybersecurity measures, and a proactive approach to stay ahead of emerging threats.

## REFERENCES

1. J. Liu et al., "Artificial Intelligence in the 21st Century," in *IEEE Access*, vol. 6, pp. 34403-34421, doi: 10.1109/ACCESS.2018.2819688, 2021.
2. J. Jha et al., "Artificial Intelligence and Applications," 1st International Conference on Intelligent Computing and Research Trends (ICRT), Roorkee, India, pp. 1-4, doi: 10.1109/ICRT57042.2023.10146698, 2023.
3. M. F. Ansari, A. Panigrahi, G. Jakka, A. Pati and K. Bhattacharya, "Prevention of Phishing attacks using AI Algorithm," 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), Bhubaneswar, India, pp. 1-5, doi: 10.1109/ODICON54453.2022.10010185, 2022.
4. S. Yang, X. Bi, J. Xiao and J. Xia, "A Text-to-Image Generation Method Based on Multiattention Depth Residual Generation Adversarial Network," 7th International Conference on Computer and Communications (ICCC), Chengdu, China, pp. 1817-1821, doi: 10.1109/ICCC54389.2021.9674427, 2021.
5. M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga and M. Bennamoun, "Text to Image Synthesis for Improved Image Captioning," in *IEEE Access*, vol. 9, pp. 64918-64928, doi: 10.1109/ACCESS.2021.3075579, 2021.
6. M. Osugi and D. V. Vargas, "Image Generation from Text and Segmentation," Tenth International Symposium on Computing and Networking Workshops (CANDARW), Himeji, Japan, pp. 206-211, doi: 10.1109/CANDARW57323.2022.00041, 2022.
7. I. Hasan, S. Rizvi, S. Jain and S. Huria, "The AI enabled Chatbot Framework for Intelligent Citizen-Government Interaction for Delivery of Services," 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 601-606, 2021.
8. S. Guefrachi et al., "Deep learning based DeepFake video detection," International Conference on Smart Computing and Application (ICSCA), Hail, Saudi Arabia, pp. 1-8, doi: 10.1109/ICSCA57840.2023.10087584, 2023.
9. B. Zhang and T. Sim, "Localizing Fake Segments in Speech," 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, pp. 3224-3230, doi: 10.1109/ICPR56361.2022.9956134, 2022.
10. Bao, Yang and Hilary, Gilles and Ke, Bin, "Artificial Intelligence and Fraud Detection Innovative Technology at the interface of Finance and Operations", Springer Series in Supply Chain Management, forthcoming, Springer Nature, <http://dx.doi.org/10.2139/ssrn.3738618>, 2020.
11. G. J. Priya and S. Saradha, "Fraud Detection and Prevention Using Machine Learning Algorithms: A Review," 7th International Conference on Electrical Energy Systems (ICEES), Chennai, India, pp. 564-568, doi: 10.1109/ICEES51510.2021.9383631, 2021.
12. S. K. Hashemi, S. L. Mirtaheri and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," in *IEEE Access*, vol. 11, pp. 3034-3043, doi: 10.1109/ACCESS.2022.3232287, 2023.
13. N. Sugunraj, A. R. Ramchandra and P. Ranganathan, "Cyber Fraud Economics, Scam Types, and Potential Measures to Protect U.S. Seniors: A Short Review," IEEE International Conference on Electro Information Technology (eIT), Mankato, MN, USA, pp. 623-627, doi: 10.1109/eIT53891.2022.9813960, 2022.

## Biometric Authentication & its Security Purposes

**Prof.S.B.Bele, & Prof. K.P. Raghuvanshi**

Assistant Professor, Department of MCA, Vidya Bharati Mahavidyalaya, Amravati, India

### **Abstract-**

Trusted user authentication is becoming an increasingly important function in a web-enabled world. The effect of an unsecure authentication system in a corporate or enterprise environment can be prosperous and can include dropping of confidential information, rejection of service, and compromise of data integrity. The value of trusted user authentication is not limited to computer or network access. Many other applications in daily life also require user authentication, such as banking, e-commerce, and can benefit from physical access control and enhanced security to computer resources.

**Keywords:** Biometric, Pattern, Iris, Authentication, Security, Sensors

### **I. INTRODUCTION**

Password less authentication plays an important role in improving this situation. By leveraging the unique physical characteristics of individuals to establish their identity, it provides unparalleled security. It locks sensitive, critical information behind the scenes of your fingerprints, iris patterns, facial features, voice patterns, and behavioral patterns like keystroke dynamics. Biometric systems can be deployed in applications ranging from physical authorization and time attendance systems to mobile devices and online transactions. This compatibility enables organizations to apply tighter security measures across multiple touchpoints, protecting sensitive data and assets.

### **II. APPLICATIONS OF BIOMETRIC AUTHENTICATION**

#### *A) Lawful Applications*

**Justice and Law Enforcement:** Biometric technology and law enforcement have a long history, and many important variations in identity management have arrived from this beneficial relationship. Biometrics implemented by the police force today is truly multimodal. Fingerprint, face and voice recognition play a unique role in improving public safety and tracking the people we are looking for.

#### *B) Government Applications*

**Border Control and Airports:** A major area of application for biometric technology is at the boundary. Biometric technology helps automate the boundary crossing process. Reliable and automated passenger screening initiatives and automated SAS help simplify the international passenger travel experience while improving the efficiency of government agencies and keeping borders more secure than ever.

#### *C) Health Care Applications*

In the healthcare sector, biometrics presents an enhanced model. Medical records are one of the most valuable personal documents; Doctors need to access them quickly and accurately. Lack of privacy and good computing can make the difference between timely and error-free detection and health fraud.

#### *D) Commercial Applications*

**Privacy:-**

As connectivity spreads across the globe, it's clear that old security methods aren't strong enough to protect what matters most. Fortunately, biometric technology is more accessible than ever, poised to provide added security and convenience for everything from car doors to phone PINs that need to be protected.

**Finance: -**

Biometric technology is widely used in finance to enhance security and convenience. By using unique biometric characteristics like fingerprints, iris, voice, and face, customers can securely access their financial data. These biometric modalities, used alone or in combination, help protect against fraud and ensure that the person accessing the account is the authorized user.

**Eyes Movement Applications: -**

Eye movement tracking applications have various uses in different industries. In the automotive industry, tracking a driver's eye movements can help measure sleepiness or drowsiness. Screen navigation applications use eye tracking to assist people with disabilities in scrolling web pages or performing actions on computers or mobile devices. In aviation, eye and head movement tracking in flight simulators can analyze pilot behavior and serve as a training tool for new pilots.

**E) Screen Navigation**

Screen navigation applications that track eye movements are indeed crucial for people with disabilities. By using cameras, these applications enable individuals to scroll web pages, write text, and perform actions on computers or mobile devices simply by clicking on buttons. This technology has been gaining significant attention due to the rapid development and the increasing demand for new methods of screen navigation, particularly on mobile devices platforms. It's exciting to see how this innovation is improving accessibility for individuals with disabilities.

**III. SECURITY NEEDS FOR BIOMETRIC AUTHENTICATION**

According to research papers, security is crucial for biometric authentication due to the following reasons: Non-repudiation: Biometric authentication provides a unique and personal identifier for individuals, making it difficult to deny their actions or presence. Difficult to replicate: Biometric traits, such as fingerprints or iris patterns, are difficult to replicate, making it challenging for unauthorized individuals to gain access. Enhanced protection: Biometric data is stored in encrypted form, adding an extra layer of protection against unauthorized access. Reduced reliance on passwords: Biometric authentication reduces the reliance on traditional passwords, which can be easily forgotten, guessed, or stolen.

**IV. SIMPLE BIOMETRIC SYSTEM ARCHITECTURE**  
**SENSOR**

The sensor is the first block of the biometric system which gathered all the crucial data for biometrics. It is the interface between the system and the natural world. Basically, it is an image acquisition system, but it also depends on the peculiarity or characteristics required that it has to be restored or not.

**PRE-PROCESSING**

The second block in a biometric system performs pre-processing tasks. Its function is to increase the input and cancel artifacts from the sensor, background noise, etc. It also performs

some kind of normalization to prepare the data for further analysis. It is the second block that executes all the pre-rectifying. Its function is to increase the input and to cancel artifacts from the sensor, background noise, etc. It performs some kind of normalization.

### **FEATURE EXTRATOR**

The third step in a biometric system is indeed the most important one. It involves extracting features to identify them later on. The goal of a characteristic extractor is to characterize an object using calculation for recognition.

### **TEMPLATE GENERATION**

The template generator plays a crucial role in the biometric system. It generates templates using the extracted features for authentication. These templates can be in the form of a vector of numbers or an image with distinct characteristics. They are stored in the database for differentiates and serve as input for similar.

### **MATCHER**

The matching phase involves using a matcher to compare the acquired template with the stored templates. Various algorithms like Hamming distance are used for this comparison. Once the inputs are matched, the results are generated

### **APPLICATION DEVICE**

An application device is a device that utilizes the results of a biometric system. Examples of such devices include the Iris recognition system and facial recognition system. They make use of biometric data for identification and authentication purposes.

### **RESEARCH**

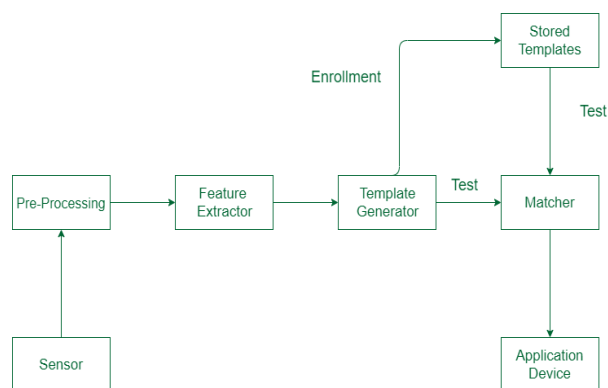
In biometric authentication, research methodology involves conducting studies to understand and improve the accuracy and reliability of biometric systems. Researchers start by selecting a specific biometric modality, such as fingerprint, iris, or face recognition. They collect a large dataset of biometric samples from individuals to create a training set. Next, researchers develop algorithms and models to extract unique features from the biometric samples. These features are then used to create templates or reference points for each individual. The templates are stored securely in a database.

#### **Data overload & accuracy**

Data overload and accuracy are important considerations in biometric authentication. When it comes to data overload, it refers to the situation where a large amount of biometric data is collected and processed. This can pose challenges in terms of storage, processing power, and efficiency. It's crucial to strike a balance between collecting enough data for accurate identification and authentication, while also considering the limitations of the system.

#### **How Biometric authentication help in Security purposes.**

- Biometric authentication enhances security by using unique physical or behavioral traits for identification.
- These traits, such as fingerprints, iris patterns, or facial features, are difficult to replicate, making it harder for unauthorized individuals to gain access.
- Biometric data is more secure than traditional methods like passwords, as it is inherently tied to the individual and cannot be easily forgotten or stolen.



Biometric System Architecture

## ResearchMethodology

In biometric authentication, researchers follow a systematic research methodology. They start by selecting a specific biometric modality, like fingerprint or face recognition. Then, they collect a dataset of biometric samples and develop algorithms to extract unique features. These features are used to create templates for each individual. Researchers evaluate the system's performance using testing protocols and metrics like False Acceptance Rate and False Rejection Rate. Statistical analysis is done to analyze the results and improve the system.

## Discussion

Biometric authentication is a fascinating field that offers secure and convenient ways to verify one's identity. It utilizes unique physical or behavioral characteristics, such as fingerprints, iris patterns, or facial features, to authenticate individuals. This technology has numerous applications, from unlocking smartphones to accessing secure facilities. It provides a higher level of security compared to traditional methods like passwords or PINs, as biometric traits are difficult to forge or replicate. However, it's important to address privacy concerns and ensure that biometric data is securely stored and used ethically. Overall, biometric authentication is an exciting field with promising advancements in enhancing security and user experience.

## FutureScope

The future scope of biometric authentication looks promising. Advancements in technology will likely lead to more accurate and reliable biometric systems. We can expect improvements in areas such as multi-modal biometrics, where multiple biometric traits are combined for enhanced security. Additionally, research and development will focus on addressing challenges like spoofing attacks and ensuring the privacy and security of biometric data.

## V. RESULT

Biometric authentication/verification provides secure and convenient identity verification using unique physical or behavioral characteristics. It offers a higher level of security differentiated to traditional methods like passwords. Advancements in technology continue to improve biometric systems, making them more accurate and reliable..

## VI. CONCLUSION

Biometric authentication is a secure and convenient method of verifying one's identity using unique physical or behavioral characteristics. It offers a higher level of security differentiated to traditional methods like passwords. Advancements in technology continue to improve biometric systems, making them more accurate and reliable. Biometric authentication has found applications in various industries, providing secure access to sensitive information and facilities. It is a promising field with ongoing innovation and potential for widespread adoption.

## REFERENCES

- [1] <https://www.seminaronly.com/computer%20science/Biometrics%20Based%20Authentication%20Problem.php>
- [2] <https://www.intechopen.com/chapters/65920>
- [3] <https://www.geeksforgeeks.org/what-is-biometrics/>
- [4] [https://www.researchgate.net/publication/46189709\\_Biometric\\_Authentication\\_A\\_Review](https://www.researchgate.net/publication/46189709_Biometric_Authentication_A_Review)
- [5] [https://link.springer.com/chapter/10.1007/1-84628-064-8\\_1](https://link.springer.com/chapter/10.1007/1-84628-064-8_1)
- [6] <https://www.mastercard.com/news/perspectives/2022/brazil-biometric-verification/?cmp=202>
- [7] <https://www.ijert.org/review-paper-on-biometric-authentication>

## Importance of Cyber Security and Cyber Security Tips

**Mrs. Rekha N. Yeotikar**

M.C.A, M.C.M. ,Dept of Management, Vidya Bharati Mahavidyalaya, Amravati

Email: rekhatilyeotikar@gmail.com

### *Abstract*

Cybercrime is one of the major crimes done by computer expert. In this paper, need of cyber security is mentioned and some of the impacts of the cybercrime Cyber security is combination of processes, technologies and practices. The objective of cyber security is to protect program, applications, network, computer and data from attack. This need is even more apparent as systems and applications are being distributed and accessed via an insecure network, such as the Internet. The Internet itself has become critical for governments, companies, financial institutions, and millions of everyday users. Networks of computers support a multitude of activities whose loss would all but cripple these organizations. As a consequence, cyber security issues have become national security issues. Protecting the Internet is a difficult task. Cyber security can be obtained only through systematic development; it cannot be achieved through hazard seat-of-the-pants methods. Applying software engineering techniques to the problem is a step in the right direction. In this paper author introduces security and privacy in online networking. Approaches to prevent, detect, and respond to cyber-attacks are also discussed.

***Keywords: Cyber security, cryptographic, cyber-attack, insecure network, systematic development, network-based software***

### INTRODUCTION

Cybersecurity is the practice of protecting systems, networks, and programs from digital attacks. These cyberattacks are usually aimed at accessing, changing, or destroying sensitive information; extorting money from users via ransomware; or interrupting normal business processes.

A major contributor to internet is that many computer systems and software applications were not designed with enough attention to security. For instance, the Domain Name System (DNS) was not designed to be completely secure.

Implementing cyber security has software, hardware and human components. Humans must implement policies such as using strong passwords and not revealing them, software must be kept up to date with patches that fix its vulnerabilities. Antivirus software and firewalls can help prevent unauthorized access to private data.

### CYBER SECURITY AND PRIVACY IN ONLINE SOCIAL NETWORK

Now a days, the idea of communication has used mainly Smartphone and computers the internet is used. Due to facing the problem of security, various cyber-crimes events happened in the past decade. Cyber security plays an important role in the current development of information technology and services. Cyber security is tried to secure the users to keep their personal and professional information undamaged from the attacks on the internet.

The significance role of cyber security is to protect networks, computers, programs from unauthorized access and loss. Most of the users are not aware of the risks and share their



information unknowingly and their lack of knowledge makes them vulnerable to cyber-attacks. So cyber security is the main concern in today's world of computing.

Social media has become very important in the life for many people. But, as with anything else online, it's important to be aware of the risks. We use social media to keep in touch with others, plan events, share our photos and comment on current events. But, as with anything else online, it's important to be aware of the risks. This are some advice on how you can keep your social media accounts safe and secure.

#### **A. Look after your logins**

One of the very achievable things of social media is that person is connected from anywhere and always. However, it's very important that where and how you can log in to your accounts our system or device. Avoid your mobile phone safe and secure

#### **B. Use strong passwords on your accounts**

We can secure ourself by using strong and unique passwords to maintain for social media accounts is one of the easiest ways to keep them secure.

### **Types of Cyber Security**

#### **1. Phishing**

Phishing is the rehearsal of distribution fake communications that look like emails from dependable sources. The goal is to bargain thoughtful data comparable to credit card details and login data. It's the greatest kind of cyber-attack. You can help defend manually over learning or an expertise solution that sieves malicious electronic mail.

#### **2. Ransomware**

It is a type of malicious software. It is considered to extract currency by blocking contact to records or the PC system until the deal is paid. Paying the ransom does not assurance that the records will be recuperated or the system returned.

#### **3. Malware**

It is a type of software intended to gain illegal right to use or to cause impairment to a system. Social engineering It is a tactic that opponents use to pretend you into illuminating delicate information. They can importune a monetarist payment or improvement access to your reserved informations. Social engineering can be collective with some of the pressures registered above to style you additional probable to connect on links, transfer malware, or belief a malicious cause. Goals The majority of the business operations run on the internet exposing their data and resources to various cyber threats. Since the data and system resources are the pillars upon which the organization operates, it drives lacking maxim that a risk to these individuals is definitely a threat to the group itself. A threat can be anywhere between a minor bug in a code to a complex cloud hijacking liability. Risk assessment and estimation of the cost of reconstruction help the organization to stay prepared and to look ahead for potential losses. Thus knowing and formulating the objectives of cybersecurity exact to every organization is crucial in protecting the valuable data. Cybersecurity is a practice formulated for the safeguard of complex data on the internet and on devices safeguarding them from attack, destruction, or unauthorized access. The goal of cybersecurity is to ensure a risk-free and secure environment for keeping the data, network and devices guarded against cyber terrorisations.

### **Goals of Cyber Security**

The definitive objective of cybersecurity is to defend the data from actuality stolen or co-operated. To attain this we aspect at 3 important goals of cybersecurity.

1. Defensive the Privacy of Information
2. Conserving the Integrity of Information

### 3. Controlling the Obtainability of information only to approved users

These objectives practice the confidentiality, integrity, availability (CIA) triad, the base of entirely safety agendas. This CIA triad model is a safety model that is intended to guide strategies for data security inside the places of a society or corporation. This model is similarly mentioned to in place of the AIC (Availability, Integrity, and Confidentiality) triad to side-step the mistake with the Central Intelligence Agency. The rudiments of the triad are reflected the three greatest vital mechanisms of safety. The CIA standards are one that greatest of the societies and businesses practice once they have connected a new request, makes a record or when assuring access to approximately information. On behalf of data to be totally safe, all of these safe keeping areas must originate into result. These are safe keeping strategies that all effort together, and hence it can be incorrect to supervise one policy. CIA triad is the greatest collective standard to measure, choice and appliance the proper safety panels to condense risk.

#### 1) Confidentiality

Making guaranteed that your complex statistics is reachable to accredited users and safeguarding no informations is revealed to unintended ones. In case, your key is private and will not be shared who power adventure it which ultimately hampers Confidentiality.

Methods to safeguard Confidentiality:

- Data encryption
- Two or Multifactor verification
- Confirming Biometrics

#### 2) Integrity

Make sure all your data is precise; dependable and it must not be changed in the show from one fact to another. Integrity ensure methods:

- No illegal shall have entrance to delete the records, which breaks privacy also. So, there shall be
- Operator Contact Controls.
- Appropriate backups need to be obtainable to return proximately.
- Version supervisory must be nearby to check the log who has changed.

#### 3) Availability

Every time the operator has demanded a resource for a portion of statistics there shall not be any bout notices like as Denial of Service (DoS). Entirely the evidence has to be obtainable. For example, a website is in the hands of attacker's resultant in the DoS so there hampers the obtainability.

## CYBER SECURITY TIPS

Cybercrime is undoubtedly one of the fastest-growing crimes in the world and it continues to impact businesses in all industries. Unless you want your company or firm's name to end up in the headlines as a result of a security breach, you need to be aware of the most up-to-date cybersecurity tips and best practices.

Staying protected from cyberattacks is challenging, however. It's difficult to keep up when cybercriminals are persistently looking for new ways to expose security risks.

Still, there are a number of cybersecurity tips that will help you prevent cyber attacks.

### 1. Keep software up-to-date

Software companies typically provide software updates for 3 reasons: to add new features, fix known bugs, and upgrade security.

Always update to the latest version of your software to protect yourself from new or existing security vulnerabilities.

### 2. Avoid opening suspicious emails

If an email looks suspicious, don't open it because it might be a phishing scam. Someone might be impersonating another individual or company to gain access to your personal information. Sometimes the emails may also include attachments or links that can infect your devices.

### **3. Keep hardware up-to-date**

Outdated computer hardware may not support the most recent software security upgrades. Additionally, old hardware makes it slower to respond to cyber-attacks if they happen. Make sure to use computer hardware that's more up-to-date.

### **4. Use a secure file-sharing solution to encrypt data**

If you regularly share confidential information, you absolutely need to start using a secure file-sharing solution. Regular email is not meant for exchanging sensitive documents, because if the emails are intercepted, unauthorized users will have access to your precious data.

On the other hand, using a secure file-sharing solution like TitanFile will automatically encrypt sensitive files so that you don't have to worry about a data breach. Your files are only as secure as the tools you chose to share them with.

### **5. Use anti-virus and anti-malware**

As long as you're connected to the web, it's impossible to have complete and total protection from malware. However, you can significantly reduce your vulnerability by ensuring you have an anti-virus and at least one anti-malware installed on your computers.

### **6. Use a VPN to privatize your connections**

For a more secure and privatized network, use a virtual private network (VPN). It'll encrypt your connection and protect your private information, even from your internet service provider.

### **7. Check links before you click**

Links can easily be disguised as something they're not so it's best to double check before you click on a hyperlink. On most browsers, you can see the target URL by hovering over the link. Do this to check links before you click on them.

### **8. Don't be lazy with your passwords**

Put more effort into creating your passwords. You can use a tool like [howsecureismypassword.net](http://howsecureismypassword.net) to find out how secure your passwords are.

### **9. Disable Bluetooth when you don't need it**

Devices can be [hacked via Bluetooth](#) and subsequently your private information can be stolen. If there's no reason to have your Bluetooth on, turn it off!

### **10. Enable 2-Factor Authentication**

Many platforms now allow you to enable 2-factor authentication to keep your accounts more secure. It's another layer of protection that helps verify that it's actually you who is accessing your account and not someone who's unauthorized. Enable this security feature when you can.

### **11. Remove adware from your machines**

Adware collects information about you to serve you more targeted ads. It's best to rid your computer of all forms of adware to maintain your privacy. Use [AdwCleaner](#) to clean adware and unwanted programs from your computer.

### **12. Double-check for HTTPS on websites**

When you're on a website that isn't using HTTPS, there's no guarantee that the transfer of information between you and the site's server is secure. Double-check that a site's using HTTPS before you give away personal or private information.

### **13. Don't store important information in non-secure places**

When storing information online, you want to keep it in a location that can't be accessed by unauthorized users.

### **14. Scan external storage devices for viruses**

---

External storage devices are just as prone to malware as internal storage devices. If you connect an infected external device to your computer, the malware can spread. Always scan external devices for malware before accessing them.

#### **15. Avoid using public networks**

When you connect to a public network, you're sharing the network with everyone who is also connected. Any information you send or retrieve on the network is vulnerable. Stay away from public networks or use a VPN when you're connected to one.

#### **16. Avoid the "secure enough" mentality**

Unless you're completely isolated from the rest of the world, there's no such thing as being "secure enough." Big companies like Facebook invest a fortune into security every year but are still affected by cyber attacks.

#### **17. Invest in security upgrades**

Following the previous tip, try to invest in security upgrades when they're available. It's better to eat the costs of security than pay for the consequences of a security breach!

#### **18. Back up important data**

Important data can be lost as a result of a security breach. To make sure you're prepared to restore data once it's lost, you should ensure your important information is backed up frequently on the cloud or a local storage device.

#### **19. Train employees**

The key to making cybersecurity work is to make sure your employees well trained, in sync, and consistently exercising security practices. Sometimes, one mistake from an improperly trained employee can cause an entire security system to crumble.

#### **20. Use HTTPS on your website**

Having an SSL certificate installed and HTTPS enabled on your website will help encrypt all information that travels between a visitor's browser and your web server.

#### **21. Employ a "White Hat" hacker**

Not all hackers are bad. Some hackers expose security risks for the sake of helping others improve their cybersecurity by keeping them aware of security flaws and patching them. These hackers are known as "white hat" hackers. It might benefit you to hire one to help you find risks you never knew you had.

### **CONCLUSION**

Any intelligence device that can pass data to one or more other devices (either through a network or not) is encompassed within the scope of cyber security that include pretty much the entire foundation of modern society. All need to be aware of cyber security as well as cybercrimes and its little seriousness about security regarding online, social and other activities through which probability of risk is higher. It causes loss of data, modifying data, removing useful information as personal details, passwords of mail accounts or bank accounts. People may also know about laws against cybercrimes or cyber laws and action which will be taken and how to fight against crimes.

### **REFERENCES**

- <https://www.getgds.com/resources/blog/cybersecurity/6-cybersecurity-threats-to-watch-out-for-in-2021>
- <https://cltc.berkeley.edu/scenario-back-matter/>
- <https://www.bitdegree.org/tutorials/what-is-cyber-security/>
- <https://www.google.com>
- Martin , john rice . cyber crimes understanding and addressing the concern of stake holders computer and security in computational and applies science

## A Survey on Blockchain Technology Concepts, Applications and Security Issues

**Jaykumar Meshram**  
meshramjaykumar@gmail.com  
Narsamma ACS College  
Amravati

**Dr. Dinesh Satange**  
dineshnsatange@rediffmail.com  
Narsamma ACS College  
Amravati

**Dr. Swapnil Deshpande**  
swapnildeshpande33@gmail.com  
S.S. Maniar College  
Nagpur

### Abstract:

Blockchain is one of the newest technologies in recent years and has become increasingly common in our daily lives. Blockchain is a decentralized, traceable, freedom, justice, transferability, authentication, anonymity, auditability and transparency tamper-proof and reliable distributed database system run by many nodes. Blockchain is used not only in crypto currencies or electronic cash, but also in other applications such as finance, healthcare, energy, reputation, insurance, IoT, manufacturing, education, and promises to provide increasingly powerful intelligence. Over the past few years, many publications and media have reported on the purpose, collaboration, development and use of blockchain.

This article aims to provide users and researchers with an overview of block chain technology security issues also focusing on applications used in block chain, concepts and related phenomena, provides detailed information about security measures.

**Keywords** blockchain, security, encryption, cryptocurrency, bitcoin, cybersecurity, vulnerability.

### Introduction:

The concept of blockchain requires controlled data storage and sharing of information and transactions in a network. The use of this technology creates a transparent environment that allows cryptographically secure transactions to be verified and confirmed by all users.

Although blockchain has seen extensive development in many areas, it is still related to some security issues and vulnerabilities caused by different types of blockchain networks, leading to problems in interactions with the blockchain.

In blockchain, information is stored in a distributed ledger. Blockchain technology ensures integrity and availability by allowing participants in the blockchain network to record, read and verify transactions recorded in a decentralized ledger. Blockchain technology also requires consensus as a law so that all participants can access common information, which is essentially a set of rules that all participants must comply with as an international agreement within the region. In a trustless environment, blockchain offers users decentralization, independence, fairness, non-interference, authentication, crime prevention, etc [1].

Blockchain is based on a distributed, immutable database that makes it easy to record assets and track transactions across a network of partners. Assets can be tangible or intangible. In a blockchain network, any value can be stored and exchanged, reducing risk and increasing efficiency for all users. Generally speaking, a blockchain is a digital record of recorded transactions. It is decentralized and not controlled by any individual, group or company. Blockchain as a technology can be difficult to change without the consent of those who use it. Blockchain stores information as a list of transactions. Network participants can read, write, and verify changes. Transactions cannot be modified or deleted. Digital signatures, hash functions, and other cryptographic functions are used to support and protect blockchain systems. This important process ensures that changes recorded in the registry are securely

protected and verified. This system is called blockchain because new blocks are connected to old blocks to form a chain. In recent years, blockchain technology has attracted the attention of academia and industry due to its advanced features. It can be used for many applications beyond cryptocurrencies. Blockchain technology has become the leading technology in networking with the Internet of Things (IoT) [2].

### Overview of Blockchain History

Blockchain was first called Bitcoin in 2009. As a public blockchain, Bitcoin is the first trusted peer-to-peer electronic cash with approximately 28.5 million electronic wallets. Since then, many other electronic currencies have been launched and wallets have increased many times. Blockchain enables the exchange of value (i.e. business) without the need for permission from a central authority. These changes are stored in a ledger maintained by a group of interconnected computers called peer-to-peer computers, rather than by a central organization like bank records. The BC system performs authentication (i.e. approval) before transaction confirmation, which plays an important role in ensuring security. The feature of BC is that its design does not require trust, and security and reliability can be provided with special mathematical or programming codes. As of 2016, most blockchain networks are used for cryptocurrency trading. Recently, the use of blockchain has gone beyond cryptocurrencies and includes IoT, artificial intelligence, etc. It started to be used in other application areas as well [3].

Chaum was the first to announce a blockchain-like system in 1982. In 1991, Haber and Stornetta used cryptography to define blockchain security. In 1993, Bayer et al. Incorporate Merkle trees into the design. In 1998, Szabo created "Bit Gold", a decentralized digital currency mechanism. In 2008, Satoshi Nakamoto introduced Bitcoin, electronic cash with a completely peer-to-peer network. Also in 2008, the term blockchain was first incorporated into the decentralized ledger behind Bitcoin transactions. In 2013, Buterin proposed the white form of Ethereum. In 2014, Ethereum generated a large number of users and on July 30, 2015, the Ethereum network became operational. The emergence of Ethereum means the birth of Blockchain 2.0 because, unlike all blockchain projects that focus on the creation of altcoins (other currencies similar to Bitcoin), Ethereum allows people to rely on printing on applications. In other words, Bitcoin is designed for distributed data, while Ethereum is designed for data storage and smart contracts (i.e. small computers). Ethereum 2.0 updates the Ethereum network and aims to increase the speed, robustness, efficiency and security of the network. The development consists of three phases from 2020 to 2022. In 2015, the Linux Foundation announced the Hyperledger project, open source software for blockchain [4].

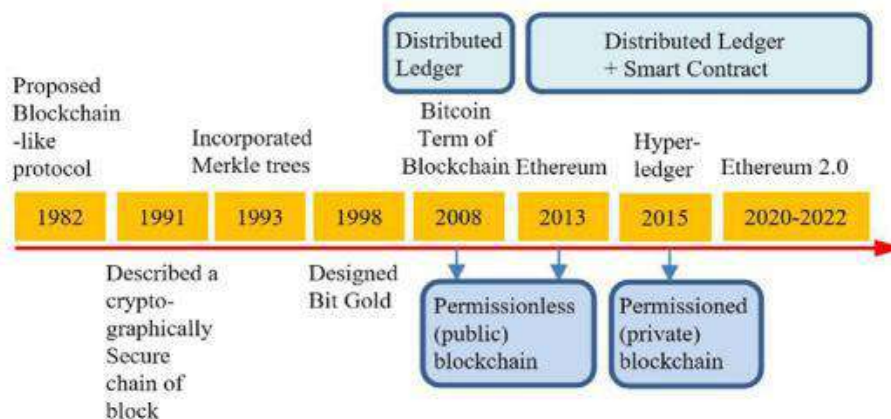
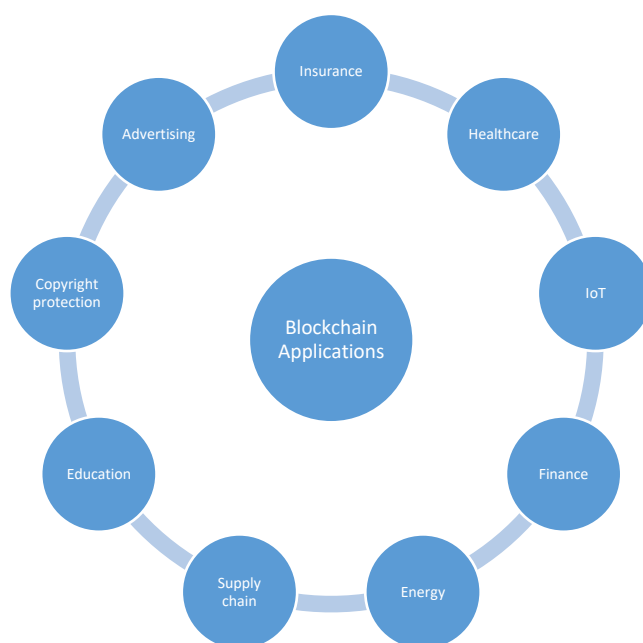


Fig. 1. History of blockchain.

**Blockchain Applications** According to an analysis, blockchain applications include cryptocurrency, finance (stock exchange, financial services, P2P financial transactions, social services, etc.), Internet of Things (IoT) (private security, e-commerce, etc.), reputation (online communities, education, etc.), security and privacy (security, risk management, privacy protection, etc.), healthcare, insurance, legal protection, energy, social applications (blockchain music, blockchain government) , advertising, national defense, mobile technology, supply chain, automobile, agriculture, self-governance, voting, education, law, asset tracking, digital information, access discovery, digital ownership management, real estate registration, etc [5].



**Fig 2 Applications of Blockchain**

## LITRATURE REVIEW

Blockchain technology has attracted widespread attention due to its versatility and potential impact. It has emerged as the technology behind Bitcoin, the cryptocurrency that first experienced a Hollywood-style boom and then crash, but is now used by other businesses as well. One of the key reasons why blockchain is so popular is its design; It uses a peer-to-peer network and records to store transactions and is a register. Digital names are stored in linked groups or blocks. Each block is cryptographically locked with the previous block and cannot be changed when a new block is added. Blockchain could be a revolutionary concept. It directly affects different businesses. But blockchain is not risk-free. These risks are mainly related to technology, implementation, investment, legal, operational, security, financial and other issues directly or indirectly related to blockchain. Information on the internet is not secure. Data is the most complete and important part of any system and will be protected with maximum security. Centralized data storage opens a window for hackers to be vulnerable to cybercrime and illegal use. Even though they have the best cybersecurity systems and the ability to solve these problems, they are not yet ready to overcome the new and clever cyber attacks developed by hackers. Therefore, strengthening the cybersecurity approach is often the most important thing for all businesses and organizations going forward. Blockchains are often replicated and shared across a network of computer systems on the blockchain. All participants in the blockchain have access to information or updates on all transactions. This information is known as Distributed Ledger Technology (DLT). Transactions on the blockchain are stored as hashes, which are cryptographic symbols that cannot be changed. Hashing is just a special algorithm. This shows that it is an immutable system with high data security. If a block in the chain

---

changes, it will be very obvious. It can be difficult for a hacker who is ready to force his way through every distribution of the chain without changing every block [6].

**Method:**

In this study, data was collected through data analysis and the obtained data was explained. Document review involves identifying documents that contain information about a case or cases under investigation. Data is collected by making explanations from keywords determined by various data collection methods. The collected data is divided into specific and related topics. The main purpose of descriptive analysis is to make its content understandable and digestible for the reader. The researcher read, organized, and digitized the data according to previously developed themes.

Focusing on blockchain security issues, this article reviews number of research papers. The main study in this review article is about the concept of blockchain, security issues and blockchain issues. This review will contribute to new research on blockchain and security.

In this article, we will first take a closer look at blockchain technology, specifically its history, cryptography are presented. Finally, this article discusses open issues and future research related to blockchainIoT systems.

First, we cover basic concepts such as blockchain, blockchain security, and search for publication and article information on the internet. IEEE Symposium, IEEE Transactions Journal etc. We analyze blockchain-related information published at leading security conferences and journals such as. In this way, we searched as many documents as possible to eliminate bias in research and conclusions.

The first step in searching for publications and information on the Internet is to identify terms such as blockchain, and blockchain security. The second way is to review information published in leading blockchain-related conferences and journals.

Blockchain is a distributed system where security is important to its success. However, despite their popularity and adoption, there is still no formal framework for examining blockchain-related threats. To fill this gap, the main goal of our work is to improve and expand knowledge on blockchain security and privacy and contribute to modeling in the field.

According to the nature of business and user needs, blockchain can be divided into private chain and public chain.

Three Aspects of Blockchain Security: The complex process of blockchain innovation raises some concerns regarding its implementation and security.

1. Privacy: Blockchain provides many functions to ensure user privacy. User keys are the only interface between users and their data. However, these keys can easily be made anonymous. Some networks use zero-knowledge authentication to ensure user privacy. Therefore, blockchain allows users to have a high level of privacy at launch and offers rich tracking opportunities.

2. Data Integrity: Blockchains are organized as records where each block is linked to adjacent blocks using a cryptographic hash function. This way, once a transaction is recorded on the blockchain it cannot be modified or deleted. Any changes made to the existing record will be processed as a routine transaction.

3. Availability: A large number of nodes make the blockchain versatile, even if some nodes are inaccessible. Since every hub in the network has a copy of the decentralized data, even in the event of a compromise, other peers can still access the actual blockchain [7].

**Conclusion:**

Blockchain technology has attracted much attention due to its diverse applications and potential impact. One of the most important reasons why Blockchain is so popular is its structural features. It uses a peer-to-peer network and records transactions that are saved as computer



files and stored in linked groups or blocks. Each block is cryptographically locked with the previous block and cannot be changed once a block is included.

This research provides an in-depth overview of blockchain technology, cryptography. A brief historical overview of blockchain is presented. Challenges and research models for creating more scalable and secure blockchain systems for large-scale deployments are presented. Finally, we hope that through our efforts, someone will gain a deeper understanding of blockchain technology. We also hope everyone pays more attention to blockchain security. The future of blockchain research is also promising in many applications. One of the main research topics is Bitcoin, as it is used every day in the cryptocurrency market. Some more areas of blockchain technology include: healthcare, public sector, blockchain as a service (BaaS), Internet of Things, cognitive power, large applications, smart deals and agreements.

### References:

1. Sujit Biswas, Kashif Shaif, Fan Li, Boubakr Nour, and Yu Wang, "A Scalable Blockchain Framework for Secure Transactions in IoT", IEEE Internet of Things Journal, Volume: 6, Issue: 3, pp. 4650 – 4659, DOI: 10.1109/JIOT.2018.2874095, June 2019.
2. Huaqun Guo Xingjie Yu, "A survey on blockchain technology and its security", Blockchain: Research and Applications, 2022.
3. Md Rafiqul Islam, Muhammad Mahbubur Rahman, Md Mahmud, "A Review on Blockchain Security Issues and Challenges", IEEE 12th Control and System Graduate Research Colloquium (ICSGRC 2021), 7 August 2021.
4. Asma Mubark Alqahtani, Abdulmohsen Algarni, "A Survey on Blockchain Technology Concepts, Applications and Security", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 2, 2023.
5. Sanjay S, Praveen S Kamath, "Blockchain Vulnerability and Cybersecurity", IJRTI | Volume 6, Issue 11 | ISSN: 2456-3315, 2021.
6. SAURABH SINGH, A.S.M. SANWAR HOSEN, and BYUNGUN YOON, "Blockchain Security Attacks, Challenges, and Solutions for the Future Distributed IoT Network", IEEE Access (Volume: 9), DOI: 10.1109/ACCESS.2021.3051602, 14 January 2021.
7. Xiaoqi Lia, Peng Jiang, Ting Chen, Xiapu Luo, , Qiaoyan Wen, "A Survey on the Security of Blockchain Systems", Future Generation Computer Systems DOI:10.1016/j.future.2017.08.020, 2017.

## A survey on face detection techniques using deep learning

**Prof.Amit B.Rehapade**

Research Scholar, S.G.B. Amravati University, Maharashtra, India.  
Email:amitrehapade@gmail.com

**Dr P. E. Ajmire**

Head. Dept. Of Comp. Science G. S. Sci., Arts & Comm. College, Khamgaon.  
peajmire@rediffmail.com

**Kiran H. Varma**

Research Scholar, S.G.B. Amravati University, Maharashtra, India.  
arow.kiran@gmail.com

### Abstract

Face detection is the first step in recognizing the facial emotions of a human being. It attracted many researchers due to its wide range of applications such as video supervision, face recognition, facial expression analysis, and object tracking. Face detection focuses on three stages Image preprocessing, feature extraction, and classification. The first step is image preprocessing. It is the process of extracting regions from images or real-time web cameras, which then act as face or non-face candidate images. Secondly, feature extraction involves segmenting the desired features from preprocessed images. Lastly, classification is a process of clustering extracted features based on certain criteria. Deep learning is and still is a buzzword these days Machine learning is considered to be the new era that provides training Computers in finding a large sample of Data and primarily describes learning at multiple levels A representation that helps make sense of data Includes text, sound, and images. In general, there are various face detection methods in deep learning which are -Viola and Jones, Conventional Neural Network (CNN), Region-based conventional Neural Network (RCNN) which is a kind of CNN extension for solving the object detection tasks, MTCNN (Multi-task convolution neural network), YOLO (You Only Look Once), YOLOv3 etc. Face detection is one of the most challenging problems of pattern recognition. Various face-related applications like face verification, facial recognition, clustering of faces, etc. are a part of face detection. Viola-Jones was designed for frontal faces, A convolutional neural network (CNN) is well suited to analyzing visual data, Multi-task Cascaded Convolutional Networks (MTCNN) is a framework developed as a solution for both face detection and face alignment, YOLO ("You Only Look Once") and YOLOv3 (You Only Look Once, Version 3) are a real-time object detection algorithm that identifies specific objects in videos, live feeds or images.

**Keywords:** Face Detection; Haar features; Skin Colour Segmentation; Viola and Jones, YOLO, CNN, MTCNN.

### Introduction.

Face detection is a fundamental problem in computer vision and pattern recognition, which has been widely studied over the past few decades. Face detection is one of the important key steps towards many subsequent face-related applications, such as face verification, face recognition, face clustering [1], etc. In recent years, deep learning methods, especially deep convolutional neural networks (CNN), have achieved remarkable success in various computer vision tasks, ranging from image classification to object detection and semantic segmentation.[1]

MTCNN is a multitask neural network model for face detection. To take into account, the performance and accuracy, and avoid the huge performance consumption caused by traditional

ideas such as sliding windows and classifiers, it first uses the small model to generate a target region candidate box with certain possibilities and then uses a more complex model for fine classification and higher precision region box regression, and makes this step recursive to form a three-layer network, namely p-net, RNet, o-net, to achieve fast and efficient face detection.[2], YOLO is a state-of-the-art deep-learning framework for real-time object detection. It is an improved model than the region-based detector and outperformed standard detection datasets like PASCAL VOC and COCO datasets [3]. Deep Learning affects the process of face detection in low-resolution surveillance videos because is common to see surveillance cameras installed in higher locations for monitoring a full urban zone and this is where the previous works on face detection have limitations due to the size of faces in which they performed. Experiments performed using our proposed model with the Deep Learning technique have shown significant improvements in the accuracy of face detection in low-resolution videos [4]. Region-based CNN (R-CNN), Fast R-CNN, Faster R-CNN [1], MTCNN[2] and YOLO [3] are popular object detection networks in recent years

## 2. REVIEW OF FACE DETECTION METHODS

2.1 Face detection using deep learning: An improved faster RCNN approach, Elsevier B.V., 15 March 2018, this paper proposed a new method for face detection using deep learning techniques. Specifically, extended the state-of-the-art Faster RCNN framework for generic object detection, and proposed several effective strategies for improving the Faster RCNN algorithm for resolving face detection tasks, including feature concatenation, multi-scale training, hard negative mining, and proper configuration of anchor sizes for RPN. Here conducted an extensive set of experiments on the well-known FDDB testbed for face detection benchmark, and achieved state-of-the-art results which ranked the best among all the published methods.



Fig 1. FDDI Images

2.2 Research on Face Detection Technology Based on MTCNN, IEEE, 2020, This paper mainly studies the application of MTCNN in face detection. Through targeted learning have a deep understanding of the hot field of computer vision, and mastered the basic principles of a variety of target detection algorithms. Through the implementation of the algorithm understand the core idea of face detection tasks and the problems to be solved in the future. the research on the detection speed of MTCNN can increase its availability in tasks with high real-time requirements. The output of MTCNN can be decomposed into three parts: face classification, face candidate frame regression, and landmark location, among which face classification and face candidate frame regression are the main contents of algorithm prediction.



**Fig2. Wider face**

This data set was provided by the Chinese University of Hong Kong. They selected 61 event categories of wider and randomly selected 40%, 10% and 50% for each category as training, verification and test sets. It contains 32203 images and 393703 faces, which vary greatly in scale, pose and occlusion. The evaluation method of MTCNN detection algorithm on test set is the same as Pascal VOC database[2].

**2.3 A Deep Learning Approach for Face Detection using YOLO**, IEEE, 27 June 2019, YOLO is a state-of-the-art deep learning framework for real-time object detection. It is an improved model then the region-based detector and outperformed on standard detection datasets like PASCAL VOC and COCO [3] dataset. Detecting the object on real-time basis is comparatively faster with respect to other detection networks. It can be concluded that processing a huge amount of data using deep learning requires a high configuration NVIDIA graphics card (GPU). If the configuration of the GPU is high, then computation of the task can be achieved at a faster rate. The accuracy of the proposed model was compared with other face detection algorithms after fine-tuning all parameters and hyperparameters of the proposed model. It was shown that proposed model accuracy was higher than the haar cascade algorithm and R-CNN based face detection model.

**2.4 Small Face Detection Using Deep Learning on Surveillance Videos**, International Journal of Machine Learning and Computing, April 2019, pp 2. This study, explored the use of Deep Learning to improve the face recognition rates in low-resolution scenarios. The results showed a significant increase in the accuracy of the proposed model with a low rate of false positives on low-resolution videos. Our results showed lower accuracy and false-positive rates when the proposed model was used on the Caviar database. In the UCSP database, the results showed an improvement of 32% in the accuracy rate. This result is far better because it improves in 10% the obtained results with the proposal without Deep Learning techniques.

**2.5 Survey of Face Detection on Low-quality Images**, IEEE, 19 Apr 2018, pp1-6., this paper, made a survey on face detection algorithms, and evaluated the representatives of them: Haar-like

Adaboost cascade and HoG-SVM as traditional methods, and faster R-CNN and S3FD as deep learning methods on low-quality images. We tested the performance degradation of the above models while changing the blur, noise or contrast level. The experiment results both hand-crafted and deeply learned features are quite sensitive to low-quality inputs. And compared to scale in variant structure, scale-variant design of neural network extracting features from multiple layers could benefit the detection of blurry tiny faces. The dataset we utilize to evaluate is the benchmark FDDB.

#### **Other Face Detection Methods**

The Viola-jones face detection technique popularly known as the Haar cascade uses edges and lines to detect faces in images or real-time videos. The feature extraction is done by computing the Haar value from the image. The darker areas are marked as one and the lighter areas with

a value of zero. The detection of an edge or line is estimated by a change in the intensities of the pixel. The objective is to find the difference between the summation of image pixels in a darker area and the summation of pixels in a lighter area. If the value is close to one edge is detected else there is no edge. In the proposed method face is detected by face mesh. The Blaze model detects the face landmarks in the image and a face mesh is formed. If only a part of the face is visible in the image, the model is efficient in reconstructing the complete face using face key points and detecting the face. The model detects and recognizes the faces in various illumination, and non-frontal images efficiently which other existing algorithms fail.

### 3. FINDINGS

The accuracy of various methods in face detection algorithms after fine-tuning all parameters and hyperparameters shown in table 1. It is shown that proposed methods accuracy varies using different datasets. model which is depicted in Fig. 3.

Method	Dataset	Accuracy
Faster RCCN	FDDB, WIDER FACE	89.6
MTCNN	WIDER FACE, Pascal VOC	85.7%
YELO	FDDB	92.2
CNN	Caviar and UCSP	92%

Table 1: Methods, Datasets and accuracy

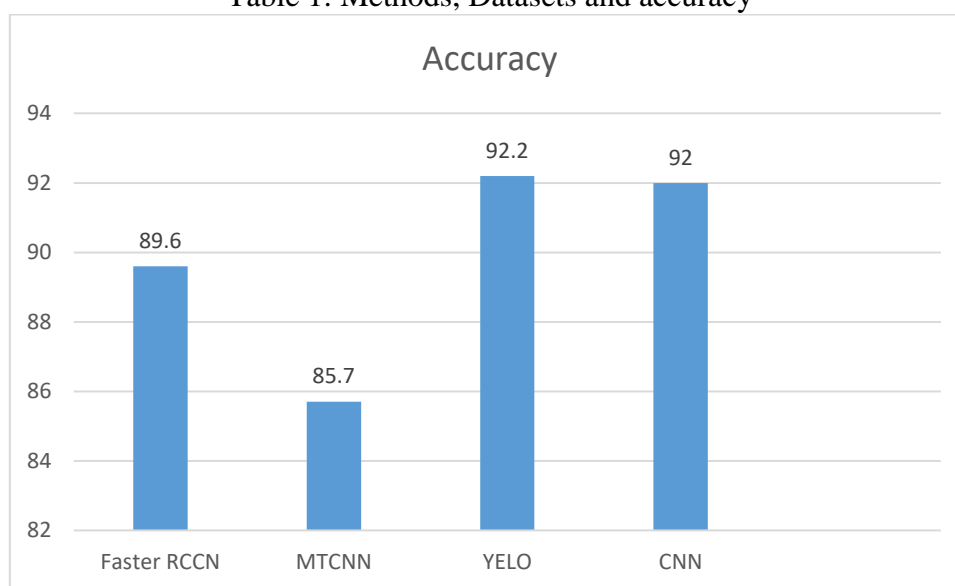


Fig. 3 Comparison of accuracy of various face detection algorithms

### 4. CONCLUSION

This paper studies the various face detection methods using deep learning to find out the best method for face detection. The deep convolutional neural networks (CNN), has achieved remarkable successes in various computer vision tasks, ranging from image classification to object detection and semantic segmentation with more than 92% accuracy. MTCNN is a multitask neural network model for face detection with 85.7% accuracy. ELO and RCCN algorithms are used for real time video face detection. The accuracy rate of YELO method is more than all the other methods but it is not convenient for small face detection method.

**REFERENCES:-**

- [1] Xudong Sun , Pengcheng Wu , StevenC.H. Hoi,” Face detection using deep learning: An improved faster RCNN approach”, Elsevier B.V., 15 March 2018,pp. 42–50.
- [2] Ning Zhang, Wuqi Gao,” Research on Face Detection Technology Based on MTCNN”, IEEE, 2020,pp 154-158.
- [3] Dweepna Garg, Parth Goel, Sharnil Pandya, Amit Ganatra, Ketan Kotecha, “A Deep Learning Approach for Face Detection using YOLO”, IEEE, 27 June 2019,pp 1-4.
- [4] Rolando J. Cárdenas, Cesar A. Beltrán, and Juan C. Gutiérrez, “Small Face Detection Using Deep Learning on Surveillance Videos”, International Journal of Machine Learning and Computing, April 2019,pp 2.
- [5] Yuqian Zhou, Ding Liu, Thomas Huang, “Survey of Face Detection on Low-quality Images”, IEEE, 19 Apr 2018,pp1-6.
- [6] Y. Wang, X. Ji, Z. Zhou, H. Wang, and Z. Li, “Detecting faces using region-based fully convolutional networks,” arXiv preprint arXiv:1709.05256, 2017.
- [7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
- [8] D. Liu, B. Cheng, Z. Wang, H. Zhang, and T. S. Huang, “Enhance visual recognition under adverse conditions via deep networks,” arXiv preprint arXiv:1712.07732, 2017.
- [9] S. V. Tathe, A. S. Narote, and S. P. Narote, “Human Face Detection and Recognition in Videos,” Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 2016, pp. 2200 – 2205.
- [10] H. Jiang and E. Learned-Miller, “Face detection with the faster r-CNN, in Automatic Face & Gesture Recognition (FG2017), IEEE, 2017,pp. 650–657.
- [11] Lin Jingdong, Wu Xinyi, Cai Yi, Yin Hongpeng. , “A survey of convolutional neural network structure optimization”. Actaautomatica Sinica, 2020, 46 (01): 24-37.
- [12] Li Peikang, Yuan Fangfang, “A brief review of target detection methods”. Journal of science and technology, 2020 (18): 157.
- [13] Li Furing. Multi feature fusion based on mtcnn for student fatigue detection [J]. Information technology, 2020, 44 (06): 108-113 + 120.
- [14] Yang Shaopeng, Liu Hongzhe, Wang Xueqiao. “Small size face detection based on feature image fusion”. Computer science, 2020,47 (06),126-132.
- [15] Shivalila Hangaragi , Tripty Singh , Neelima N, “Face Detection and Recognition Using Face Mesh and Deep Neural Network”, Elsevier B.V.,2023,714-749.

## **A Deep Dive Into Extended Reality -Six Sense Integration -Merging Real And Virtual World**

**Miss. Sidra-tul-munteha mohd Sabir**

Msc II year, Department Of Computer Science, Vidyabharati Mahavidyalaya, Amravati  
Email id: [sidratulmuntehams@gmail.com](mailto:sidratulmuntehams@gmail.com)

**Miss. Afifa gulam ahmad Saudagar**

Msc II year, Department Of Computer Science, Vidyabharati Mahavidyalaya, Amravati  
Email id: [afifasaudagar39@gmail.com](mailto:afifasaudagar39@gmail.com)

**Prof. Dipika S.Harode,**

Department Of Computer Science, Vidyabharati Mahavidyalaya, Amravati  
Email id: [dipikaharode11@gmail.com](mailto:dipikaharode11@gmail.com)

### **Abstract:**

The merging of Extended Reality (XR) and sixth sense technologies is a burgeoning field with immense potential to redefine human-computer interaction. This review aims to explore the current state of research on their integration, highlighting key findings, applications, and future directions. As Extended Reality (XR) redefines user experiences across diverse fields. Extended Reality (XR), encompassing Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR), has revolutionized user experiences across various fields. However, the next frontier lies in seamlessly integrating XR with sixth sense technologies, enabling deeper immersion and intuitive interaction through bio-sensing. This paper explores the synergistic potential of this integration, analyzing its impact on user experience, outlining promising applications, and identifying key challenges and future directions. We argue that the fusion of XR's immersive environments with sixth sense's biofeedback and contextual awareness capabilities can unlock an unprecedented level of human-computer interaction, paving the way for personalized, adaptive, and hyper-realistic experiences

### **Keywords:**

Extended reality, sixth sense, bio-sensing, haptic feedback, immersive interaction, human-computer interaction, user experience, applications, challenges, future directions

### **Introduction:**

Mixing extended reality (XR) with sixth sense technology has the potential to create incredibly immersive and intuitive experiences that blur the lines between the physical and digital worlds.

### **Extended Reality (XR):**

XR encompasses a spectrum of immersive technologies that blend real and virtual environments. It includes:

**Augmented Reality (AR):** Superimposes Digital Information onto the real world, viewed through a device like a smartphone or smart glasses. Imagine seeing repair instruction on machinery or historical facts popping up as you explore landmark.

**Virtual Reality (VR):** Creates a fully immersive, computer-generated environment experienced through a headset. Imagine attending virtual concert, collaborating in design session or undergoing therapy in simulated scenarios.

**Mixed Reality (MR):** Combines element of AR and VR, allowing real and virtual object to interact in real time, imagine manipulating 3d model in your living room or participating in holographic training exercises.

**Sixth Sense Integration:**

Six-sense refer to expanding our current five sense (sight, touch, taste, smell and hearing) with technological capabilities. This could involve:

**Biometric Sensors:** Monitoring vital signs like heart rate and brain activity to track emotion, fatigue or focus.

**Haptic Feedback:** Providing realistic textures and sensations through devices worn on the skin or clothing.

**Small and Taste Simulation:** Delivering virtual scents and flavors through specialized devices, enhancing immersive experiences.

**XR And Sixth Sense Integration:**

Integrating Six-Sense capabilities into XR environment can create even more profound and impactful experiences.

Imaging:

**VR Training simulations:** Feeling Virtual Texture of objects, experiencing heat and wind changes, and even tasting virtual food samples for enhanced skill development.

**AR-Assisted Learning:** Receiving haptic feedback while assembling furniture or feeling the emotional pluse of historical characters through holographic presentations.

**Remote Collaboration:** Having a realistic sense of presence in virtual meetings, felling handshakes virtually and sharing immersive spatial data.

**Algorithms and design for integrating XR and Six-Sense Technology:**

Integrating XR and Six-Sense Technologies is a complex task that requires a combination of various algorithms and design principles. Here's an overview of some key areas:

**Algorithms:**

**Real-Time Sensor Fusion:** Algorithms are needed to process and combine data from various XR sensors (e.g., head tracking, hand gestures, biometrics) and six-sense devices (e.g., haptic feedback, smell diffusers) in real-time. This creates a unified sensory experience that reacts dynamically to user interactions.

**Sensory Simulation:** Algorithms translate digital information into realistic sensory stimuli for smell, taste, and touch. This involves understanding human sensory perception and generating appropriate signals for six-sense devices based on virtual elements.

**Adaptive Personalization:** Algorithms can personalize the sensory experience based on individual user preferences and biofeedback data. This ensures optimal comfort and engagement for each user by adjusting intensity, type, and timing of sensory stimuli.

**Machine Learning:** Machine Learning Algorithms can be used to analyze user data and preferences, allowing the system to adapt and improve the sensory experience over time.

**Design principles:**

**User-Centered Design:** The integration should prioritize user comfort, safety, and well-being. Interfaces need to be intuitive and non-overwhelming, considering potential sensory overload and individual needs.

**Data Privacy And Security:** Robust security measures are crucial to protect sensitive user data collected from sensors and biometrics. Transparent data handling practices and user control over information are essential.

**Ethical Considerations:** Design should address potential ethical concerns like addiction, manipulation, and impact on mental health. Transparency, user agency, and clear guidelines are necessary to ensure responsible use.



**Accessibility and Inclusivity:** The system should be accessible to people with disabilities and diverse physical abilities. This requires designing for different sensory modalities and providing alternative interaction methods.

**Interoperability and Standardization:** Open standards and protocols are needed to enable seamless integration across different xr platforms and six-sense devices, fostering wider adoption and innovation.

#### **Specific technologies:**

**Haptic Technology:** Various haptic technologies exist, from vibration motors to exoskeletons, each with its advantages and limitations. Choosing the right technology depends on the desired sensory experience and application.

**Smell and Taste Simulation:** While still in early stages, technologies like scent diffusers and taste buds on the tongue are being explored to generate virtual smells and tastes.

**Brain-Computer interfaces (bcis):** Though futuristic, bci technology holds potential for directly translating brain activity into sensory experiences, bypassing traditional interfaces.

#### **Model of integrating XR And Six-Sense Technology:**

Here is a possible model of XR with six-sense technology:

##### **Hardware:**

**Head-mounted display (HMD):** This is the primary display that users will see through. It should be high-resolution and have a wide field of view to create a truly immersive experience.

**Haptic gloves:** These gloves will provide users with a sense of touch in the virtual world. They can be used to feel objects, textures, and even other people.

**Olfactory display:** This device will emit smells that correspond to what users are seeing and doing in the virtual world.

**Gustatory display:** This device will allow users to taste things in the virtual world. It is still in the early stages of development, but it has the potential to revolutionize the way we experience virtual reality.

**Biometric sensors:** These sensors will track users' heart rate, respiration, and other physiological responses. This information can be used to personalize the experience and make it more realistic.

##### **Software:**

**Rendering engine:** This software will create the graphics and sounds that users see and hear in the virtual world. It should be able to handle the complex demands of six-sense technology.

**Physics engine:** This software will simulate the physics of the virtual world, so that objects move and interact realistically.

**Sensory feedback software:** This software will translate data from the haptic gloves, olfactory display, gustatory display, and biometric sensors into sensory experiences for the user.

#### **Challenges in XR with Six-Sense Integration:**

While XR And Six-Sense Technologies hold immense potential, integrating them seamlessly comes with several challenges:

##### **Technical:**

**Hardware limitations:** Current devices are bulky, expensive, and have limited sensory fidelity. Creating miniature, comfortable, and affordable devices that deliver realistic touch, smell, and taste sensations remains a hurdle.

**Real-time processing:** Integrating and processing data from multiple sensors and six-sense technologies in real-time to create a synchronized and responsive experience requires significant processing power and efficient algorithms.

#### **Sensory fidelity and personalization:**

**Individual variability:** Sensory perception varies greatly between people. Creating personalized experiences that cater to individual differences in sensitivity and preferences adds complexity.

**Limited understanding of senses:** Our Scientific understanding of how humans perceive smell, taste, and touch is still evolving, making it difficult to accurately replicate these sensations in a virtual environment.

#### **Privacy and security:**

**Biometric data collection:** Integrating Six-Sense Technologies often involves collecting sensitive biometric data. Ensuring user privacy and data security in these immersive environments is crucial.

**Virtual embodiment manipulation:** The ability to manipulate users' virtual avatars and sensory experiences raises ethical concerns about potential misuse and manipulation.

#### **Ethical**

#### **considerations:**

**Impact on Mental Health:** The potential impact of extended exposure to highly immersive and sensory-rich XR environments on mental health and well-being needs careful consideration.

**Accessibility and Equity:** Ensuring Equitable access to these technologies for people with disabilities and diverse socioeconomic backgrounds is vital to avoid exacerbating existing inequalities.

#### **Social and societal impact:**

**Addiction and Escapism:** the immersive nature of XR raises concerns about potential addiction and escapism, particularly with the addition of multisensory experiences.

**Real-World Impact:** The blurring of lines between the physical and virtual worlds has societal implications in areas like social interaction, work-life balance, and perception of reality.

#### **Applications:**

**Gaming:** Imagine being able to feel the weight of a gun in your hand, smell the gunpowder, and even taste the blood of your enemies. Six-sense technology could take gaming to a whole new level of immersion.

**Education:** Students could learn about history by walking through the streets of ancient rome, or about biology by dissecting a virtual frog. Six-sense technology could make education more engaging and interactive than ever before.

**Training:** Soldiers could train for combat in a virtual battlefield that feels and smells like the real thing. Doctors could practice surgery on virtual patients. Six-sense technology could revolutionize the way we train for a variety of professions.

**Therapy:** People with phobias could be exposed to their fears in a safe and controlled environment. People with chronic pain could experience pain relief through virtual reality. Six-sense technology could have a major impact on mental and physical health.

**Conclusion:**

Integrating XR and six-sense technologies is a dynamic field with continuous advancements. Utilizing powerful algorithms and adhering to ethical design principles is crucial for creating immersive, responsible, and beneficial experiences for all users. Collaboration between researchers, developers, and ethicists is key to unlocking the full potential of this exciting technology. Despite these challenges, ongoing research and development efforts are addressing these issues. With careful consideration of ethical and societal implications, XR-Six-Sense Integration has the potential to revolutionize various aspects of life, offering exciting possibilities for education, healthcare, entertainment, and more. It's important to approach this technology responsibly and collaboratively to ensure its benefits reach everyone in a positive and sustainable way."

**References:****Research Papers :**

- [1] "Crafting the Senses: Designing for Extended Reality and the Internet of Things" by Pranav Mistry and Patricio Freire (2017):  
<https://www.jetir.org/papers/JETIR1701013.pdf>
- [2] "Sensory Augmentation for Mixed Reality: State of the Art, User Acceptance, and Future Directions" by Tobias Knodel, Michael Naceri, and Matthias Rauter (2019):  
<https://www.nature.com/articles/srep42197>
- [3] "Six Senses of Presence: VR beyond Vision and Hearing" by Jeremy Bailenson, Nick Yee, Matei Zaharia, and Daniel Nowosielski (2007):  
<https://ieeexplore.ieee.org/document/10179147/>
- [4] "Haptic Interfaces for Immersive Virtual Reality" by Eunyoung Kim, Minyoung Kim, Sunghyun Kim, and Byungjin Kim (2020):  
<https://www.sciencedirect.com/science/article/abs/pii/B9780128217504000116>
- [5] "The 6 senses of the future: How VR is evolving beyond sight and sound" by The Next Web: <https://thenextweb.com/conference/themes>
- [6] "The Future of Smell and Taste in VR" by XR Today:  
<https://www.tandfonline.com/doi/full/10.1080/17458927.2018.1556952>
- [7] "The Challenges and Opportunities of XR and Six-Sense Integration" by Futurism:  
<https://futurism.com/>
- [8] "The Ethical Implications of Six-Sense Technology" by MIT Technology Review:  
<https://www.technologyreview.com/magazines/the-ethics-issue/>

**Books:**

- [9] "The Augmented Mind: How VR Will Revolutionize Reality" by Chris Milk (2020)
- [10] "Reality+: Virtual Worlds and the Future of Experience" by David J. Chalmers (2020)
- [11] "Beyond Vision: The Hidden Senses That Shape Our World" by Michael J. Slepian (2020)

## Cyber Security: Hacking, Child Pornography, Virus Dissmination

**Miss. Gayatri Dilip Wange**

Final year PG Student  
Department of Computer Science  
Vidyabharati Mahavidyalaya,  
Amravati- 444706  
[gayatriwange340@gmail.com](mailto:gayatriwange340@gmail.com)

**Miss. Diksha Laxman Pandey**

Final year PG Student  
Department of Computer Science  
Vidyabharati Mahavidyalaya,  
Amravati- 444706  
[dikshapandey2912002@gmail.com](mailto:dikshapandey2912002@gmail.com)

**Miss. Mayuri Arun Deshmukh**

Dept. of Computer Science  
Vidyabharti Mahavidyalaya,  
Amravati - 444706  
[deshmukhamayuri@gmail.com](mailto:deshmukhamayuri@gmail.com)

### ABSTRACT

Cyber Security plays an important role in the field of information technology. Securing the information have become one of the biggest challenges in the present day. Whenever we think about the cyber security the first thing that comes to our mind is ‘cyber crimes’ which are increasing immensely day by day. Various Governments and companies are taking many measures in order to prevent these cybcrimes. Besides various measures cyber security is still a very big concern to many. This paper mainly focuses on challenges faced by cyber security on the latest technologies .It also focuses on latest about the cyber security techniques, ethics and the trends changing the face of cyber security.

**Keywords:** cyber security, cybercrime, cyber ethics, social media, cloud computing, android apps.

### INTRODUCTION

Today man is able to send and receive any form of data may be an e-mail or an audio or video just by the click of a button but did he ever think how securely his data id being transmitted or sent to the other person safely without any leakage of information?? The answerlies in cyber security. Today Internetis the fastest growing infrastructure in every daylife. In today’s technical environment many latest technologies are changing the face of the man kind. But due to these emerging technologies we are unable to safeguard our private information in a very effective way and hence these day cybercrimes are increasing day by day. Today more than 60 percent of total commercial transactions are done online, so this field required a high quality of security for transparent and best transactions. Hence cyber security has become a latest issue. The scope of cyber security is not just limited to securing the information inIT industry but also to various other fields like cyber space etc.

### CYBER SECURITY

*Cyber Security is the body of technologies, processes, and practices designed to protect networks, devices, programs, and data from attack, theft, damage, modification or unauthorized access.* Privacy and security of the data will always be top securitymeasures that any organization takes care. We are presently living in a world where all the information is maintained in a digital or a cyber form. Social networking sites provide a space where users feel safe as they interact with friends and family. In the case of home users, cyber-criminals would continue to target social media sites to steal personal data. Not only social networking but also during bank transactions a person must take all the required security measures.

**CYBER SECURITY:** - Cyber Security involves protection of sensitive personal and business information through prevention, detection and response to different online attacks. Cyber security actually preventing the attacks, cyber security.

**PRIVACY POLICY:** - Before submitting your name, e-mail, address, on a website look for the sites privacy policy.

**KEEP SOFTWARE UP TO DATE:** - If the seller reduces patches for the software operating system your device, install them as soon as possible. Instilling them will prevent attackers from being able to take advantage. Use good password which will be difficult for thieves to guess. Do not choose option that allows your computer to remember your passwords.



## TYPES OF CYBER CRIME

- **HACKING:** - Hacking in simple terms means an illegal intrusion into a computer system and network. It is also known as cracking.
- **DENIAL OF SERVICE ATTACK:** - This is an act by the criminals who floods the bandwidth of the victim's networks or fills his E-mail box with spam mail depriving him of the service he is entitled to access or provide.
- **CHILD PORNOGRAPHY:** - The internet is being highly used by its abuses to reach and abuse children sexually, worldwide. As more homes have access to internet, more children would be using the internet and more are the chances of falling victim to the aggression of Pedophiles.
- **VIRUS DISSEMINATION:** - Malicious software that attaches itself to other software VIRUS, WORMS, TROJAN HORSE.

- **COMPUTER VANDALISM:** - Damaging or destroying data rather than stealing or misusing them is called cyber vandalism. These are programs that attach themselves to a file and then circulate.
- **CYBER TERRORISM:** - Terrorist attacks on the internet is by disturbed denial of service attacks, hate websites and hate E-mails, attacks on service network etc.

### **Advantages of the cyber security**

#### **a) Data safety from hackers**

- Cyber security is designed to reduce the chance of data breaches against criminals. It uses tools and techniques like the DLP technique in conjunction with firewalls, web servers, and access control methods for protection. It also restricts resource access based on user tasks and powers or network connections.

#### **b) Reduces computer crash**

- While working with technology, the user must deal with various harmful attacks that may result in freezing screens and computer crashes. This can bring the work life of people working with tight deadlines at risk. These kinds of problems can be diminished by cyber security and lower the hindrance of working with technology.

#### **c) Decreased data theft hazard**

- The major benefit of cyber security is that it prevents unauthorized or malicious user access to the system. The high-security protocol is implemented to protect against major data theft and makes the experience a lot more relieving.

#### **d) System availability and improved data**

- If a system is free from threats due to cyber security, it can boost the effectiveness of data and its network. It also improves the quality of data as it is less harmful.

#### **e) Protect business reputation**

- Every organization's primary strategy is to win customers' trust, but a data breach can weaken the whole effort and bond of trust. Various examples have proved that data breaches have badly spoiled the business reputation because, after an attack, they failed to get the customer retention needed to strengthen brand loyalty. Organizations use technologies like network security and cloud security to avoid these sudden setbacks in the system and strengthen the security, which can also open new paths to future recommendations, ventures and expansions.

#### **f) Assist remote working**

- Cyber securities always protect analytics, strategies, and sensitive data that risk being leaked or hacked. Rather some organizations or business uses multiple remote models for their workflows. Still, it became more popular after COVID-19, where 80% of workers worked from home with their personal or professional devices, Wi-Fi, and IoT. This result in the increase of average data breach costs that make it necessary for a business to protect its sensitive data.

#### **g) Saves the bottom line**

- Cybercriminals or cyber crimes are the prime rivals of any business or individual that can suddenly take everything from bed to floor, including its sales and revenue. With low, competitive criteria, a business can't survive its continuity. Therefore cyber security has some developed technologies that defend businesses from reaching their bottom line.

#### **h) Cyber posture is improved**

- Digital protection provided by cyber security to the firms provides safety, liberty, and flexibility to the employees in freely accessing the internet. Cybersecurity technology continuously increases its safety posture by tracking all the systems with a single click.

Cyber security organizations can protect and respond during and after a cyber-attack. Cyber security protocols are strengthened to prevent threats.

### **Disadvantages of Cyber Security**

#### **a) Not affordable to everyone**

Users or businesses have to buy their services and pay for maintenance, which seems an expenditure to them. Usually, small or medium business needs more finances to protect their system and data from internal or outside cyber-attacks. They need to be aware of the advantage of using cyber security in business and invest less in cyber security. Even an individual using a system and internet couldn't afford an antivirus or firewall for their system and doesn't feel the need for it. Rather some free antivirus and window defender already installed in window help in prevention, but nothing is 100% secure.

#### **b) Can be complicated**

Cyber security measures are hard to understand for its user, normal persons, or business persons as they require a lot of time and effort. Suppose the user needs help understanding how to use cyber security, then instead of benefit. In that case, it can damage data loss, or hackers can easily take advantage of it. If a business doesn't have a proper security mechanism, it can be easily trapped and attacked by hackers with various methods. Cyber security experts must break the complexity of getting through cyber security to avoid damage.

#### **c) Security patches may backfire**

To secure the system, security experts always work on designing security patches against vulnerabilities, and once they release a new security update or patch, the hackers start their work. They try to find the weakness mended in patched files by comparing the patches and unpatched files. Then unpatched files are attacked, which is why patches can backfire on the system it was meant to secure.

#### **d) Need of constant monitoring**

As we know, hackers and cybercriminals continuously work to penetrate a business network. To tackle them, businesses have to monitor their cyber security constantly. It has two benefits. One, it keeps the system up to date, finding threats before they create harm and ensuring everything is in place.

#### **e) Slow down the system**

One of the best and most dedicated security systems consists of several passwords and checks all the system files. This can consume lots of time, resulting in slow system processing and the productivity of the person working on it.

#### **f) Can be risky**

Sometimes implementing cyber security measures can be risky for individuals or businesses because they have to compromise their data. It also increases the risk of security breaches that result in loss of money, customer trust, and the company's reputation.

### **Importance of Cybersecurity**

- Cybersecurity is just an ethical practice to protect our devices from such hackers and make them more secure. People involved in cybersecurity perform security measures and operations in order to keep our data and devices safe. Cybersecurity basically deals with protecting our network, devices, and data from illegal and unauthorized access by other people. Hackers and cybercriminals use the Internet as an opportunity to crack into other's people devices by using spyware, malware and carrying out cyber attacks.
- The main purpose of Cybersecurity is to protect all the users on the Internet from infected files, malware, and digital attacks which lead the users to access private sensitive information of users, extort ransom from users by using their private data or even disrupting important critical infrastructure like shutting down power supplies and military infrastructure.

- Cybersecurity helps to solve pre-built vulnerabilities in applications and helps them to remain stable throughout. More and more devices are getting connected to the Internet, hence it is more and more important to secure all the devices over the Internet to protect them all against unauthorized access.

## CONCLUSION

Computer security is a vast topic that is becoming more important because the world is becoming highly interconnected, with networks being used to carry out critical transactions. Cyber crime continues to diverge down different paths with each New Year that passes and so does the security of the information. The latest and disruptive technologies, along with the new cyber tools and threats that come to light each day, are challenging organizations with not only how they secure their infrastructure, but how they require new platforms and intelligence to do so. There is no perfect solution for cyber crimes but we should try our level best to minimize them in order to have a safe and secure future in cyber space.

## REFERENCE

- 1]. A Sophos Article 04.12v1.dNA, eight trends changing network security by James Lyne.
- 2]. Cyber Security: Understanding Cyber Crimes- Sunit Belapure Nina Godbole
- 3]. Computer Security Practices in Non Profit Organisations – A NetAction Report by Audrie Krause.
- 4]. A Look back on Cyber Security 2012 by Luis corróns – Panda Labs
- 5]. International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 Page nos.68 – 71 ISSN 2229-5518, “Study of Cloud Computing in HealthCare Industry “ by G.Nikhita Reddy, G.J.Ugander Reddy
- 6]. IEEE Security and Privacy Magazine – IEEECS “Safety Critical Systems – Next Generation “July/ Aug 2013.
- 7]. CIO Asia, September 3rd, H1 2013: Cyber security in malasia by Avanthi Kumar.
- 8]. M. M. Yamin, B. Katt, and V. Gkioulos, “Cyber ranges and security testbeds: Scenarios, functions, tools and architecture,” Computers and Security. 2020.
- 9]. M. A. Khan and K. Salah, “IoT security: Review, blockchain solutions, and open challenges,” Futur. Gener. Comput. Syst., 2018.
- 10]. C. Lee, H. Bin Yim, and P. H. Seong, “Development of a quantitative method for evaluating the efficacy of cyber security controls in NPPs based on intrusion tolerant concept,” Ann. Nucl. Energy, 2018.
- 11]. J. Srinivas, A. K. Das, and N. Kumar, “Government regulations in cyber security: Framework, standards and recommendations,” Futur. Gener. Comput. Syst., 2019.



## **Novel Design and Implementation of the Personalized Search Engine in Context with User Keywords Profile and Keywords Optimization Technique**

**Mr. K.P.Raghuvanshi #1**

Department of Research and PG Studies in Science and Management, MCA Programme, Vidya Bharati Mahavidyalaya, Amravati.

**Mr. S.B.Bele #2**

Department of Research and PG Studies in Science and Management, MCA Programme, Vidya Bharati Mahavidyalaya, Amravati.

### **Abstract:**

Huge amount of information are available on internet. Retrieval of this information from internet is very difficult task. Search engine play a very important role for retrieving the information from internet. The ultimate objective of search engine is to give a most related possible result to the users. Today users would like to personalize their search, to get the result precisely that satisfied the user's requirement/goal/objective. The search result that are obtained from the present high rated search engine are huge in number out of which first few are relevant to the users need, Author proposes a new personalized search engine based on user keyword profile and long tail keywords optimization technique. Keyword optimization is a vital component of the overall search engine optimization process, and proper keyword selection and keyword placement based upon thorough keyword research. Keyword optimization works well only when it is used with the right techniques in right combination. The user interest keywords based on long tail keywords optimization technique. In the present study new approach is applied to achieve better results in relevancy, recall rate, (Recall rate is the fraction of the documents that are relevant to the query that are successfully retrieved.) and high retrieval accuracy compared to the present traditional searching engine. The basic framework and the basic functions module of the system are described in this paper. For the purpose of simulation of system, IUKBPSE will be developed by using PHP platform.

**Keywords:** Information retrieval, Search Engines, user, keyword profile, keywords optimization technique and recall rate.

### **I. INTRODUCTION**

With the fast development of Internet technology, search engine has been widely used. Now a day's Internet is fast growing technology in its 4th generation. Due to the large scale use of Internet, search engine become an important technology. In order to satisfy need to personalize the search. The personalized service has become an active research area with the rapid development of search engines. To provide personalized search service for users, personalized search engine were developed. The most common personalization approaches were user behavior, Ontology, user interest mining, users' access interest, User Profile, User Interests. The present study of personalization is based on user keyword profile using long tail keyword optimization technique. Keyword optimization is an integral component of search engine optimization. It is one of the foundations of SEO. It can greatly affect the conversion rate. Keyword optimization (also known as keyword research) is the act of researching, analyzing and selecting the best keywords to target the result. Keyword search optimization is a critical step in initial stages of search engine. In the present study author proposes to design a

personalized search engine model that is based on user keyword profile & long tail keywords optimization technique. Long tail keywords are a type of keyword phrase that has at least three, and some times as many as five words in the phrase these keywords are highly specific. Choosing the right long tail keyword combinations involves careful research around a particular topic. Experiments by the method of simulation will be performed to verify the effectiveness of proposed model. Simulation tests were capable to improve the search result effectively. The propose work is aimed to improve the accuracy and relevancy of search result that will be obtained by personalized search engine user.

## II. REVIEW OF LITERATURE:

The literature survey for the personalized search engines is done regeously. Li Yong *et al* [1] design and implement personalized search engine base on shared knowledge base and mining of the users' interest from database. Ontology was use effectively to improve the recall and precision ratios. LiuZhongbao[2] personalized search engine model is also based on user interest mining. Xiang-dongChen *et al* [3] suggested that users' access interests were introduced into the design of personalized search engine by using web mining technology. Firstly, the users' access interest transactions were gained by interest algorithm via mining the users' logs. Secondly, it presents a method to compute session similarity of transactional unit and transaction and sets up an interest similarity matrix for clustering by setting the suitable threshold value. Clustering along with page rank algorithm is used to get accuracy. HaoChen *et al* [4] suggest a new personalized information retrieval model which is based on imported user interest which provides fast and accurate information to the user. XueLi *et al* [5] gives realistic a user-based personalized searching system, and provides with its general framework and implementation of system functions. According to the user's browsing history and personal information, the author builds the user-interest model based on tourism areas, and then realizes the personalized information recommendation technology and information expanding searching. Lai, J. *et al* [6] used a method base on similarity score to derive users' documents profile.

## III. METHODOLOGY

### 3.1 Search Engine Architecture

Numbers of phases as shown in Fig1 are involved in retrieval of information needed by the user. The Search Engine Architecture Figure 1.

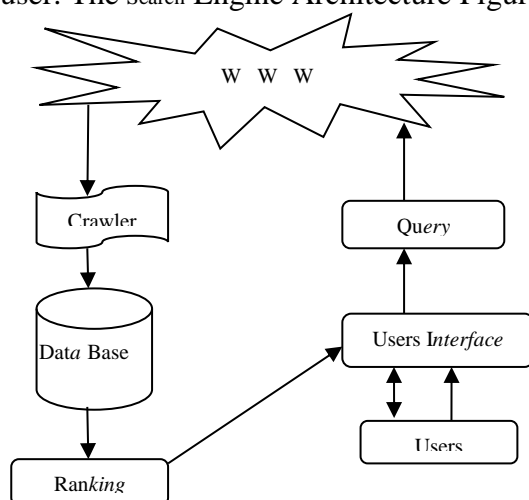


Fig: 1. Search Engine Architecture.

### ***The basic functional modules in the architecture of search engine are :***

- Users: a person who uses the search engine over internet.

- The user interface module: The user interface module is the interface which provides the registration or login interface for user. In the user interface module, the user input some keywords. Such information is called the user’s explicit personalized information which can compose the default information of the user keywords profile. These keywords are stored in to the Database.
- Query Engine: In this module users enters the keywords and send it to the search engine.
- Crawler: Crawler gets the data from the World Wide Web (WWW) and stores it into the search engine database.
- Rank: The rank of the web page is measured on the basis of click through Rate (CTR).

**A new approach for architecture of personalized search engine is proposed here which is based on User Keywords Profile and Keywords optimization technique shown in Figure 2.**

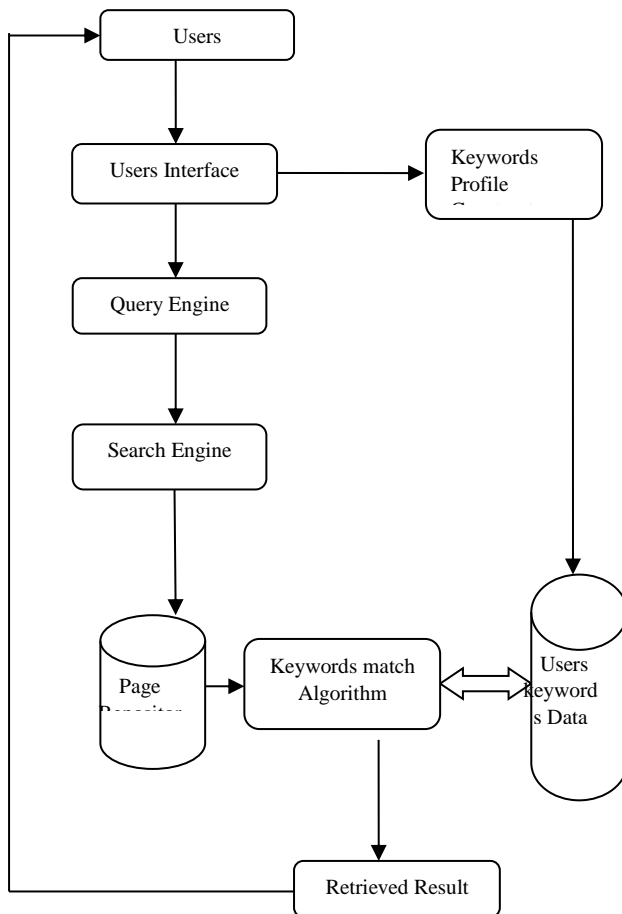


Fig: 2. The proposed Architecture of Personalized Search Engine

**The functional module in the proposed architecture of personalized search engine is given below:**

- Users: a person who uses the internet.
- The user interface module: In this module we create GUI for the users and provide the login and keyword registration forms.
- Keywords Profile Constructor: It constructs user keywords profile.

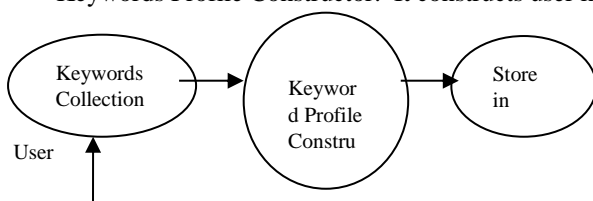


Fig 3: Design of user keywords profile constructs.

- In the Keywords Collection user enter interested keywords typically via HTML forms on which they want to search the information. Such information is called the user's explicit personalized information, which can compose the default information of the user keyword profile.
- The users profile here is a set of keywords. The profile can be extracted from the documents or user's himself can create it. The keywords profile may represent area of interest. These keywords information profile are stored in the user's keyword Database.
- Query Engine: In this module query engine accept keyword from users and then it will be forwarded to search engine for further process.
- Search engine: Search engines helps to discover billions of web sites available on the internet. They are the integral part of WWW. In this module search engine send the user keyword on internet using Crawler. Crawler is a set of program that find out appropriate data on internet. Crawler brings up to date data from WWW and store it into its database.
- Page Repository: This module collects the result from search engine and store in XML database.
- Keyword Matching Algorithm: There are many matching algorithm are available like Brute-force algorithm, Boyer-Moore algorithm [12, 18] and Knuth-Morris-Pratt algorithm [13]. It works on string only. Author propose a new keywords (character and string) matching algorithm based on User's Keyword Profile shown in figure 4.
  1. Get the keyword from user
  2. Store in User Database
  3. Get SE result
  4. Store browser data in Page Repository
  5. Initialize input character/string
  6. Match User's keyword from database with Page Repository
  7. Display filtered result based on user's profile.

Fig: 4 Search Algorithm based on User's Keyword Profile

- Retrieved Result: The results are collected and send to the users.

#### IV. CONCLUSION

The effectiveness of proposed model will be shown by stimulation method. The proposed work is aimed to improve the accuracy and relevancy of search result that will be obtained by personalized search engine users. The goal of personalized retrieval is to provide correct information within a very short period of time. The result retrieved on the basis of user keyword will generate fast, accurate and relevant result as compared to the earlier work done.

#### REFERENCES

- [1].LiYong,LiGuan-yu "Research and realization of personalized search engine based on ontology" Published in International Conference on Network and Parallel Computing Workshops,Sept 2007. Page(s): 1016-1020.
- [2]. Liu Zhongbao" Research of a personalized search engine based on user interest mining" Published in International Conference on Intelligent Computing and Integrated Systems (ICISS),Oct 2010 page(s):512-515.
- [3].Xiang-dongChen, LinHuang "The research of personalized search engine based on users' access interest" Published in CIIA on Computational Intelligence and Industrial Applications, Nov-2009. page(s):337-340.
- [4]. Hao Chen, Cheng Zeng "Personalized Information Retrieval Model Based on User Interests" Published in International Conference on Computer Science and Software Engineering, Dec-2008 .page(s):12-14.
- [5]. XueLi , JunpingDu ,LinglingZi, ShimoharaK."Study and implementation of personalized searching system based on user's interest model" Published in IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS), Oct- 2012.page(s): 1-5.
- [6]. LaiJ, SohB." Personalized Web search results with profile comparisons" Published in Third International Conference on Information Technology and Applications, july- 2005.page(s): 573-576.

- 
- [7]. Wei-Chao Li and, Jin-Guang Liu 2013 “Design and Implementation of the Personalized Search Engine Based on the Improved Behavior of User Browsing Research” Journal of Applied Sciences, Engineering and Technology 5(4): 1257-1262, 2013 ISSN: 2040-7459; e-ISSN: 2040-7467 © Maxwell Scientific Organization, 2013 Submitted: June 28, 2012 Accepted: August 08, 2012 Published: February 01, 2013
- [8]. Jiandong Cao, Yang Tang, Binbin Lou “Personalized Meta-search Engine Design and Implementation”
- [9]. Preeti Naval, Priyanka Singh “A Survey on Personalized Meta Search Engine” International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 3, March 2012 ISSN: 2277 128X
- [10]. Amit Kumar, Vishnu Sharma, Shishir Kumar “A Comparative Analysis of Various Exact String-Matching Algorithms”.
- [11]. Christian Charras, Thierry Lecroq “Exact String-Matching Algorithms” Handbook June 2004
- [12]. Robert S. Boyer, J. Strother Moore “A Fast String Searching Algorithm” In Communications of the ACM Volume 20 / Number 10 / October, 1977
- [13]. Knuth D.E., Morris (Jr) J.H., and Pratt V.R.,” Fast pattern matching in strings,” SIAM Journal on Computing 6(1): page(s): 323-350, 1977.
- [14]. Aho A.V. and Cora sick M.J.” Efficient String Matching”: An Aid To Bibliographic Search Comm. ACM, 18: page(s): 333-340, 1975
- [15]. Cyclone, NSlab, RIIT “String Matching Algorithm” PowerPoint Presentation
- [16]. Timo Raita, Tuning the Boyer-Moore-Horspool String Searching Algorithm; Software— Practice And Experience, Vol 22(10). Page: 879–884 (October 1992)
- [17]. Horspool R N. Practical fast searching in strings. Software practice and Experience, 1980, Vol 10(6): page(s): 501-506
- [18]. Lecroq T., A variation on the Boyer-Moore algorithm, Theoretical Computer Science Vol 92(1): page(s): 119—144
- .

## How biometric affects Cyber Security

**Miss. Vaishnavi Govardhanrao Dhole**

M.Sc. II year, Department Of Computer Science, Vidyabharati Mahavidyalya, Amravati  
Email id: dholevaishnavi22@gmail.com

**Miss. Sakshi Sanjay Rathi**

Msc II year, Department Of Computer Science, Vidyabharati Mahavidyalya, Amravati  
Email id: sakshirathi3111@gmail.com

**Prof. Dr. Shilpa B. Sarvaiya,**

Department Of Computer Science, Vidyabharati Mahavidyalya, Amravati  
Email id: sarvaiya.shilpa@gmail.com

### ABSTRACT

Cybersecurity is crucial to the information technology industry. One of the main problems of the modern world is information security. The first thing that springs to mind when considering cyber security is the rapidly rising number of cybercrimes. Numerous governments and businesses are implementing numerous efforts to deter these cybercrimes. Despite these precautions, many people continue to have serious concerns about cyber security. This essay primarily addresses the difficulties that modern technology present for cyber security. It also emphasizes the most recent developments in cyber security methods, morality, and fashions that are redefining the field.

### KEYWORDS

Keywords: cyber security, cyber crime, Biometrics Security.

### INTRODUCTION

Cybersecurity is the discipline of defending hardware, software, and data connected to the internet from cyberthreats such as theft, hacking, and data breaches. It includes a range of tools, procedures, and methods intended to protect devices, networks, and private data against harm or illegal access. Network security, application security, endpoint security, data security, identity management, and cloud security are important aspects of cybersecurity. As people depend more and more on digital technologies for personal and professional purposes, cybersecurity is essential to maintaining the privacy, availability, and integrity of data and systems.

Even the newest technology, such as online banking, cloud computing, mobile computing, and e-commerce, require a high level of security. These technologies now require heightened protection since they include some very sensitive personal data.

### DEFINATION

The process of protecting networks, computers, servers, mobile devices, electronic systems, and data from hostile intrusions is known as cyber security. It is often referred to as electronic information security or information technology security..

### Techniques for Cyber Security :-

#### Password security and access control

A key component of information security has always been the idea of a user name and password. This can be among the initial steps taken in terms of cyber security.

**Data authentication**

Before downloading, any papers that we receive must always be authenticated. This means that they must be verified to have come from a reputable source and to not have been altered. Usually, the anti-virus software on the devices is responsible for authenticating these documents. Therefore, to safeguard the devices against viruses, effective anti-virus software is also necessary.

**Software for Antivirus**

Computer programs known as antivirus software are designed to identify, stop, and eliminate harmful software, including worms and viruses. The majority of antivirus software comes with an auto-update capability that allows the application to download virus profiles as soon as they are found, allowing it to scan for new infections right away. Every system needs anti-virus software as a basic requirement.

**Firewalls:-**

**A firewall is a piece of hardware or software that helps block viruses, worms, and hackers from infecting your computer via the Internet.** Every message that enters or exits the internet is filtered by the firewall, which checks each one and deletes any that don't fit the predetermined security requirements. Firewalls are crucial in identifying malware because of this.

**Malware Detectors :-**

This software typically checks all of the system's files and papers for dangerous viruses or malicious code. Malicious software is generally referred to as malware and includes programs like Trojan horses, worms, and viruses.

**Cyber crime :-**

Cybercrime refers to any illicit action where the primary tool for commission and theft is a computer. The concept of cybercrime has been broadened by the U.S. Department of Justice to encompass any illicit behaviour that makes use of a computer to save evidence. The increasing number of cybercrimes includes both computer-based versions of pre-existing crimes like identity theft, stalking, bullying, and terrorism, which have become serious issues for individuals and countries, as well as crimes that have been made possible by computers, like network intrusions and the spread of computer viruses. Cybercrime is generally understood to be any crime that uses a computer and the internet to steal someone's identity, sell illegal goods, harass victims, or interfere with operations through malicious software. Cybercrimes are becoming more common nationwide as technology and healthcare executives play a bigger role in people's lives every day. Silicon Valley Bank found that these crimes will rise in tandem with technological companies' belief that cyberattacks are a serious advancement. Cyberattacks pose a threat to both their data and their business.

**Cyber****Security :-**

Cybersecurity is a collection of technologies, procedures, and practices created to guard against cyberattacks that could harm or allow illegal access to networks, computer systems, programs, and data. Security in the context of computers encompasses both physical and cyber security.

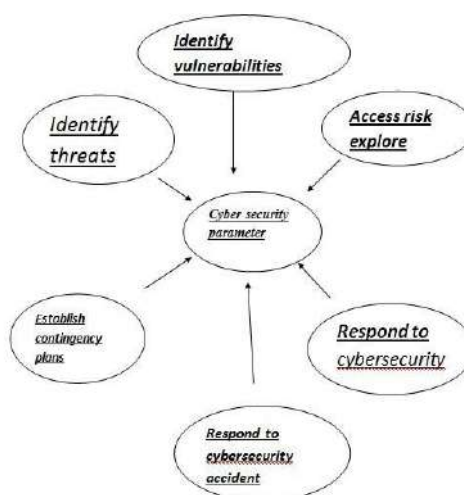
The fact that many computer systems and software programs were created with insufficient consideration for security is one of the main causes of the internet. The Domain Name System (DNS), for example, was not intended to be totally secure.

Cybersecurity implementation involves hardware, software, and human components. Policies like creating secure passwords and keeping them secret must be implemented by humans, and software needs to be updated with updates that address vulnerabilities. Firewalls and antivirus programs can aid in preventing unwanted access to personal information.

**The parameters of cyber security:-**

The following are the criteria for cyber security:

1. Determine the dangers.
2. Recognize weak points.
3. Examine access risks
4. Create an emergency plan.
5. Address a cyber security mishap.
6. Develop a backup plan.



### How to stop, identify, and react to cyber attacks:-

- A. Providing staff with cyber security training.
- B. Set up, use, and keep antispymware and antivirus software up to date on all work computers.
- C. Protect your internet connection with a firewall.
- D. As soon as software updates become available, download and install them on your system and apps.
- D. Create backup copies of crucial company documents and data
- E. Manage physical access to your PCs and network hardware
- F.
- G. Protect your wireless network. Make sure your workplace's Wi-Fi network is concealed and safe if you have one.
- H. Demand unique user accounts for every worker.
- I. Restrict employee access to information and data, as well as their ability to install software. Make frequent password changes.

### Biometrics Security:-

Consider the term "bio" in the context of "biology" and biometrics. The scientific study of life and living things is known as biology. "Metrics" is a rules-based method of measuring data that is frequently used for comparison or tracking reasons. It is not merely a tool that the rest of the world (except from the USA) uses to calculate distances between locations. Metrics are quantitative, but biology is primarily qualitative.

How does biometrics affect cybersecurity?

In recent years, there has been a significant surge in the use of biometric authentication. These days, biometric authentication isn't simply for entering extremely secure spaces. The usage of biometric authentication has expanded across applications, from simple daily use-cases like taking attendance and unlocking your phone, to entry into server rooms and safes.

Certain systems employ biometrics as one of the authentication methods, while other systems require it based on the use case and criticality. In any case, biometrics has improved security. The majority of firms choose the latter since authentication requires both something you are (biometrics) and something you know/have (passwords, authentication devices). This guarantees a person's strict identification and adds another level of protection. It therefore restricts breaches. Certain extremely secure server rooms, for instance, require both a password and facial recognition to be allowed to enter.

One of the most innovative security innovations is biometrics, which is simple to use and "difficult to break through." To bolster that claim, consider this statistic: over the past five



years, there has been a 90% growth in the use of biometrics. Therefore, it is undeniable that biometrics have become the new norm in security.

### **Risks of Biometrics in Security:-**

#### **Not impervious to hacks of personal information:-**

Biometric authentication does, without a doubt, improve security. Biometrics are not impervious to data intrusions, though. Your biometrics are obtained if a malevolent actor gains entry to the database. This creates a risk not only to the company you work for, but also to your identity because hackers may use your biometrics against you.

#### **Confidentiality:-**

Personal biometrics are traits unique to each person. Your privacy may therefore be violated if an unauthorized individual has access to your biometrics. The reason this matters most for facial biometrics is that your appearance can be used to identify you if someone gains access to the database.

#### **Inaccuracy and deception:-**

Not all biometrics use all available biometric data. They employ partial data for authentication even if they keep complete data, in order to speed up the process and allow for unforeseen little discrepancies. This indicates that just a portion of the biometric data is used by these systems. Because of this, authentication may be inaccurate, and if someone discovers the data points the system utilizes for authentication, they may be able to fraudulently circumvent it.

#### **System Errors:-**

The world we live in is not perfect. So, the possibility of anything going wrong never goes away. Failures in the biometric authentication system could be quite inconvenient. In situations where it's one of the authentication alternatives, it might not be a huge concern. For instance, you can unlock your phone with a password or facial recognition if the fingerprint scanner isn't working. However, the issue arises when a system that requires biometric authentication malfunctions. For instance, in the event that a room requires fingerprint authentication and the scanner malfunctions, you will be without alternative access until the equipment is repaired or the system is bypassed.1qq

#### **Reference :-**

1. [https://www.researchgate.net/publication/3278329\\_Biometric\\_security\\_technology](https://www.researchgate.net/publication/3278329_Biometric_security_technology)
2. <https://www.softwaresecured.com/post/risks-and-benefits-of-biometrics-in-security#what-is-the-impact-of-biometrics-on-cybersecurity>
3. Book: JainNandakumar\_BiometricAuthenticationSystemSecurityUserPrivacy\_IEEEComputer2012
4. [https://www.researchgate.net/publication/262284740\\_Teaching\\_the\\_security\\_mindset\\_with\\_reference\\_monitors](https://www.researchgate.net/publication/262284740_Teaching_the_security_mindset_with_reference_monitors)
5. <https://www.mdpi.com/2078-2489/7/2/23>
6. <https://scholars.unh.edu/cgi/viewcontent.cgi?article=2083&context=thesis>
7. [https://www.researchgate.net/publication/372483435\\_Cybercrime\\_Monitoring\\_System\\_for\\_Online\\_Security\\_Experts](https://www.researchgate.net/publication/372483435_Cybercrime_Monitoring_System_for_Online_Security_Experts)
8. AIG Study: Systemic Cyber Attacks Likely in 2017; Financial Services, Power/Energy, International Cyber Conflicts Key Concerns. (2017, May 10). Retrieved September 4, 2019, from <https://www.businesswire.com/news/home/20170510005781/en/AIG-Study-Systemic-Cyber-Attacks-2017-Financial>.
9. Martin , john rice . cyber crimes understanding and addressing the concern of stake holders computer and security.in computational and applies science ‘MIT A.
10. Research paper : A Study Of Cyber Security Challenges And Its Emergning Trends On Latest Technologies G.Nikhita Reddy1 , G.J.Ugander Reddy2
11. The Reference Monitor Concept as a Unifying Principle in Computer

## Analyzing the Challenges of Marathi Textual Data for Sentiment Analysis

**Ram B. Ghayalkar**

Asst. Prof. Shri R. L. T. College of Science Akola  
rambg29@gmail.com

**Prof. Dr. D. N. Beseekar**

Principal, Shri Shivali College of Arts, Commerce & Science, Nimba  
dnbeseekar@gmail.com

### **Abstract:**

Sentiment analysis is important tasks in Natural Language Processing. There are significant work happened in various foreign languages like English, Arabic, Russian, Mandarin, and also Indian languages such as Hindi, Bengali, Tamil. Marathi is third popular language in India but also due to its low resources language there is lot of work to be in Marathi. In this paper presents while analysing Sentiment Analysis in Marathi and its versions lot of difficulties come front so overall analysis of possible challenges and their study, also it opens many way to researcher in the field of Sentiment Analysis.

**Keywords:** NLP, Sentiment Analysis(SA), Sarcasm Detection, Word sense disambiguation (WSD)

### **Introduction:**

Sentiment analysis is computational study of emotions, opinions and mainly the sentiment expressed in the text by user. Sentiment analysis is a challenging task due to many challenges which are associated while processing natural language. Sentiment Analysis (SA) is a natural language processing task that deals with analyzing emotions, feelings, and the attitude of a speaker or a writer from a given piece of text. Sentiment Analysis involves capturing of user's behaviour, likes and dislikes of an individual from the text. The main goal behind sentiment analysis is to identify sentiment associated with the text by extracting sentimental context from the text.

There are different classification levels in SA:

**Document-level:** Document-level SA aims to classify an opinion of the whole document as expressing a positive or negative sentiment.

**Sentence-level:** Sentence-level SA aims to classify sentiment expressed in each sentence which involves identifying whether sentence is subjective or objective.

**Aspect-level:** Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities which is done by identifying the entities and their aspects.

### **Related Work:**

Despite the fact that Sentiment Analysis (SA) research has reached mainstream product scenarios, significant advances in the discipline have been accomplished from all research perspectives. Code-mixing techniques and associated studies have been around for decades. The early research attempted to determine the language in mixed code texts by learning the structure of the language from the informant and/or the given text. It was discovered that rule detection by pure text alone is unsuccessful and is heavily reliant on information containing predefined linguistic rules [1].

In [2] paper, presented MarathiSarc - a dataset of labelled Marathi tweets for sarcasm detection. Dataset contains 2400 tweets in Marathi language. Further discussed our dataset collection and

the annotation policy. Finally presented some of the baseline sarcasm detection experiments performed on dataset.

Lexical analysis tools were utilised to improve the accuracy of transliteration generated by statistical methodologies, and the results were proven by [3].

In [4] have published a monolingual Marathi corpus for un-supervised language modeling tasks. Presented MahaCorpus, MahaSent, MahaNER, and MahaHate datasets and their corresponding

MahaBERT models fine-tuned on these datasets. Their aim to move ahead

of benchmark datasets and prepare useful resources for Marathi. The re-sources are available at <https://github.com/l3cube-pune/MarathiNLP>.

In [5] paper, have presented L3CubeMahaSent the first major publicly available dataset for Marathi Sentiment Analysis which consists of ~16000 distinct tweets. Also described the annotation policy which used for manually labelling the entire dataset, performed 2class and 3class sentiment classification to provide a benchmark for future studies.

The work established the efficacy of a text normalisation strategy focused on managing abbreviations, spelling errors, missing punctuation, slang, wordplay, censor evasion, and emoticons. During a recent study on the problem, I coined the term "textual code-switching" and developed a method for identifying texts that contain code-switching [6].

Clustering algorithms can improve sentiment classification accuracy by applying strategies such as sense-based and cross-lingual word clustering by word sense [7].

Uses social media data to perform mixed-script language recognition and concludes that supervised learning outperforms dictionary-based approaches. Uses social media data to perform mixed-script language recognition and concludes that supervised learning outperforms dictionary-based approaches [8].

To demonstrate the usage of sentiment analysis techniques on transliterated content, they used bilingual dictionary approaches and HSWN for sentiment score computation, reaching 80% accuracy. Another person utilised the genetic algorithm to recognise Marathi and Sanskrit words [9].

### ***General Challenges for Sentiment Analysis***

**Tone:** Tone is difficult to translate literally and even more difficult to recognise in writing.

**Polarity:** Positive (+1) and negative (-1) polarity scores for words like "love" and "hate" are high.

**Sarcasm:** Irony and sarcasm are used in informal chats and memes on social media **Emojis:** One of the problems with text-based social media content, such as Twitter, is that it is dense with emojis. Natural Language Processing (NLP) tasks for specific languages are trained.

**Idioms:** Machine learning programmes are incapable of comprehending figures of speech. For example, the algorithm will be perplexed by the statement "not my cup of tea" because it takes things literally.

**Negations:** Negative terms such as not, never, can't, were not, and so on might cause confusion in the Machine Learning model.

**Comparative sentences:** Comparative sentences can be tricky because they don't always give opinions. A lot has to be taken from it.

**Employee bias:** staff input is crucial for developing corporate culture, boosting sales strategies, and lowering staff turnover.

**Multilingual sentiment analysis:** When languages are mixed into multilingual sentiment analysis, all of the issues stated above compound.

**Audio-Visual Data:** Videos and text data are not the same thing. The challenge is that the video must not only be transcribed, but it may also have captions that must be evaluated for brand logos. [9][10]

### **Analyzing Challenges for Sentiment Analysis in Marathi Language:**

There is large Linguistic Diversity in Marathi language because it is third large used language in India with multipledialects. It is the challenging task to analyse the text as well as sentiment in Marathi due following reasons:

- Variations in Grammar and Vocabulary
- Dialectical Differences
- Handling Code-Mixing with Other Languages

### **Contextual Information Contextual Ambiguity:**

Marathi, like any language, has phrases and expressions with ambiguous meanings. Contextual understanding becomes crucial for accurate sentiment classification, as the sentiment may vary based on the surrounding text.

Identifying the context of the text becomes an important challenge to address in SA. Behaviour/use of the word changes in a great aspect based on the context.

1. चित्रपटातील गाणी उल्लेखनीय आहेत

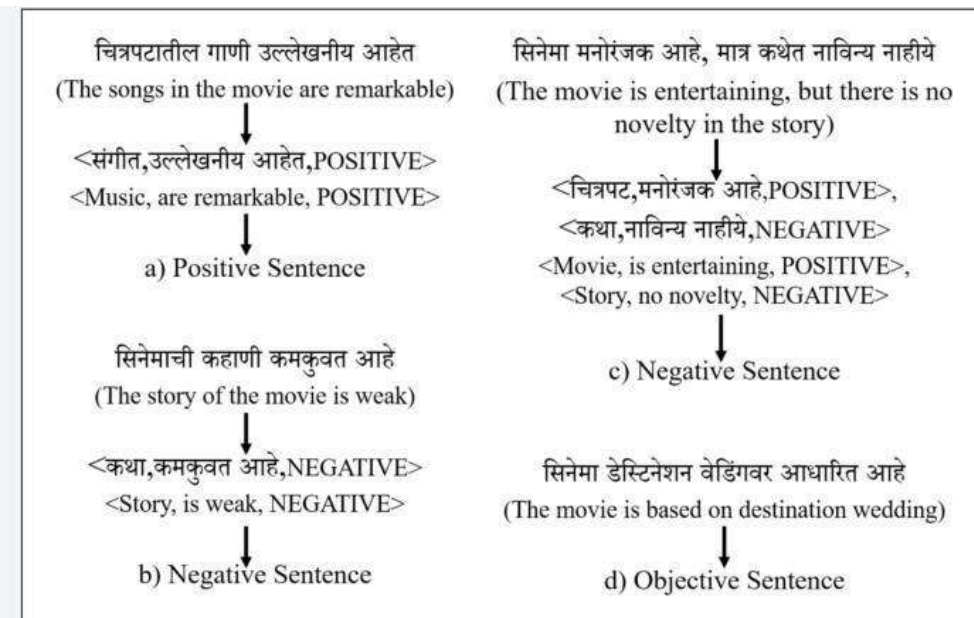
2. सिनेमा मनोरंजक आहे, मात्र कथेत नाविन्य नाहीये

3. सिनेमाची कहाणी कमकुवत आहे

Here in the above context of information indicating different meaning like sentence 1,3 indicating negative and sentence 2 indicating positive meaning.

### **Sarcasm Detection**

SARCASM detection is an important processing problem in natural language processing (NLP), which is needed for better understanding to serve as an interface for mutual communication between machines and humans. To understand this is to underline the basic problem behind it - being able to detect the contradiction.[11]



### **Word Sense Disambiguation**

Word sense disambiguation (WSD) is the problem of determining in which sense a word having a number of distinct senses is used in a given sentence. The same word can have multiple meanings, and based on the sense of its usage the polarity of the word also changes.

Statements: 1. दारूने तो ठोकला.

2. जो

ठोकला तो दारू.

Here in above sentences 1 indicating positive and 2. Indicating negative  
Sentiment analysis on Marathi textual data poses several unique challenges due to the linguistic characteristics and cultural nuances associated with the Marathi language. Below are some common challenges that researchers and practitioners may encounter in this context:

- **Linguistic Diversity:** Marathi exhibits a diverse range of linguistic features, including variations in grammar, vocabulary, and writing styles. Different dialects and colloquial expressions can make sentiment analysis more complex.
  - **Code-Mixing:** Users often mix Marathi with English or other languages in online communication. Analyzing sentiment in code-mixed text requires models that can effectively handle multilingual content and language switches.
  - **Lack of Annotated Datasets:** Building a comprehensive and well-annotated Marathi sentiment analysis dataset can be challenging. Limited labeled data for training models may hinder the development of accurate and robust sentiment analysis systems.
  - **Slang and Informal Language:** Marathi text often includes slang, colloquialisms, and informal expressions. Sentiment analysis models need to be trained to recognize and interpret these variations accurately.
  - **Cultural Sensitivity:** Sentiments in Marathi text can be deeply tied to cultural references and events. Models trained on generic sentiment analysis data may not capture these cultural nuances, leading to potential inaccuracies.
  - **Domain Specificity:** Sentiment analysis models trained on generic datasets may not perform well when applied to specific domains, such as news, social media, or customer reviews in Marathi. Domain adaptation is essential to enhance model performance.
  - **Negation and Double Negation:** The presence of negations and double negations in Marathi sentences can alter the sentiment. Models must effectively handle linguistic constructs that negate or reverse the sentiment expressed in a statement.  
**Statement:** "ह्या विषयामध्ये अनेक नकारात्मक शब्द आहेत, परंतु हे वाक्य सकारात्मक प्रभाव दर्शवते." (In this statement, there are many negative words, but the sentence reflects a positive impact.)
  - **Imbalanced Class Distribution:** Sentiment analysis datasets may exhibit imbalances in the distribution of sentiment classes. This can impact the model's ability to accurately classify sentiments, especially for less frequent classes.
  - **Dynamic Language Evolution:** Like any living language, Marathi evolves over time. Changes in language usage, the emergence of new terms, and shifts in sentiment expressions may require continuous model adaptation to maintain effectiveness.
- Addressing these challenges requires a combination of linguistic expertise, domain knowledge, and the development of robust NLP models tailored specifically for Marathi sentiment analysis. Researchers and practitioners in this field should be mindful of the intricacies of the Marathi language and its cultural context when designing and implementing sentiment analysis systems.

**Conclusion:**

Sentiment Analysis is now day pay vital role for decision making for individuals, company, organizations, government every alternative sector. In this paper analyse the several general challenges in sentiment analysis as well as specific analysis of occurring challenges while working with Marathi textual data. Also this paper opens up the future research work and new dimensions for researcher.

**References:**

- [1] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, 23(1), 11.
- [2] Pravin K. Patil and Prof. S. R. Kolhe, "MarathiSarc: A Marathi Tweets Dataset for Automatic Sarcasm Detection of Marathi Tweets", *Grenze International Journal of Engineering and Technology*, June Issue, Grenze ID: 01.GIJET.8.2.16 © Grenze Scientific Society, 2022
- [3] Ansari, M. A., & Govilkar, S. (2018). Sentiment analysis of mixed code for the transliterated hindi and marathi texts. *International Journal on Natural Language Computing (IJNLC)* Vol, 7.
- [4] Raviraj Joshi, "L3Cube-MahaNLP: Marathi Natural Language Processing Datasets, Models, and Library", *L3CubePune arXiv:2205.14728v2 [cs.CL]* 31 May 2022
- [5] Atharva Kulkarni, Meet Mandhane, Manali Likhitkar, Gayatri Kshirsagar, and Raviraj Joshi, "L3CubeMahaSent: A Marathi Tweet based Sentiment Analysis Dataset", *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 213–220 April 19, 2021. ©2021 Association for Computational Linguistics
- [6] Khan, J., & Lee, S. (2021). Enhancement of Text Analysis Using Context-Aware Normalization of Social Media Informal Text. *Applied Sciences*, 11(17), 8172.
- [7] Jardim, S., & Mora, C. (2022). Customer reviews sentiment-based analysis and clustering for market-oriented tourism services and products development or positioning. *Procedia Computer Science*, 196, 199-206.
- [8] Babu, N. V., & Kanaga, E. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, 3(1), 1-20.
- [9] Rao, L. (2022). Sentiment Analysis of English Text with Multilevel Features. *Scientific Programming*, 2022.
- [9] Ramnath Mahadeo Gaikwad, Rajashri Ganesh Kanke, Manasi Ram Baheti, "Review on Sentiment Analysis of Marathi Language of Maharashtra, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*", Volume 11 Issue VIII Aug 2023-
- [10] Mr. Ram B. Ghayalkar, Prof. Dr. D. N. Besekar, "Review on Challenges of Sentiment Analysis", *International e-Conference on New Horizons And Multidisciplinary Applications In Science And Technology In Association with International Journal of Scientific Research in Science and Technology* Volume 9 | Issue 6 | Print ISSN: 2395- 6011
- [11] Ashwitha A, Shruthi G, Shruthi H R, Makarand Upadhyaya, Abhra Pratip Ray, Manjunath T C, "Sarcasm detection in natural language processing", *ScienceDirect*, 2020 Elsevier Ltd. <https://doi.org/10.1016/j.matpr.2020.09.124>

## Biometric its application and challenges in Internet of Things

**Aloksingh M. Thakur<sup>1</sup>, Chandrakant R. Mankar<sup>2</sup>, Dr. V. M. Patil<sup>3</sup>, Dr. D.N. Satange<sup>4</sup>**

1,2 Research Scholar 3,4 Research Guide

1,4 Shrimati Narsamma Arts, Commerce & Science College Kiran Nagar, Amravati

2,3 Shri Shivaji College of Art's, Commerce & Science, Akola

**Abstract:** The biometric has been a trusted factor for authentication for years, due to its unique behavioral it is widely accepted. Nowadays it has been an important factor in the Internet of Things, as different IoT devices connected to applications used by humans are growing rapidly and from a security aspect it is accepted with strength. For researchers' biometrics and its advancement is of interest in their research. In the Internet of Things, biometrics is used in various applications like health monitoring systems, smart doors, and most commonly in mobiles. This paper gives a paradigm to the researcher to help in understanding the need for biometrics in IoT and the challenges faced, also providing the future direction of research that needs to be done. This paper includes information showing the study of biometrics, its various types available along their strength and weaknesses.

**Keywords:** Biometric, Keystroke, Iris, Fingerprint, Face Recognition, etc.

### Introduction:

[1-4] Humans are blessed with unique biometric factors by nature, these factors are widely used for authentication it's application in daily life in various like security, attendance, and E-transactions. The biometric factors are divided into two sub-types according to their behavioral and occurrences. The most widely used is the fingerprint and face which is mostly used in mobile for security purposes.

- Different types and sub-types of biometric factors:

Depending upon the uniqueness and characteristics the biometric factors are classified into two main types as shown in fig.1 , some of the commonly and widely used biometrics factors for security are shown in fig.2.

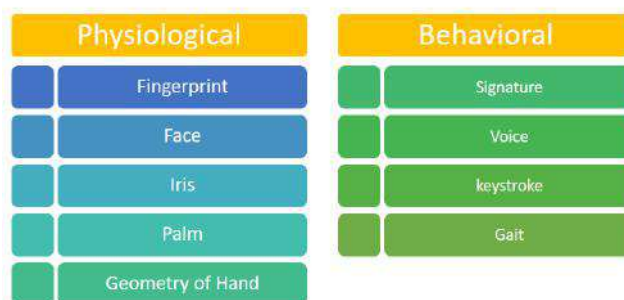


Fig.1: Classification of Biometrics Factors

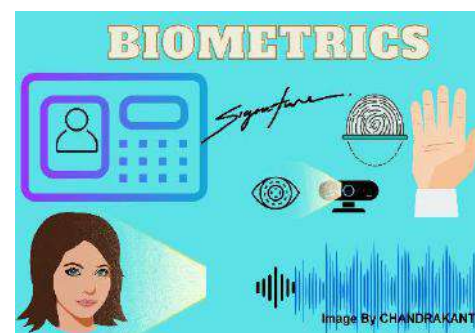


Fig. 2: Biometrics Factors

1. Fingerprint Recognition: This is the most widely used in the authentication of someone, due to the pattern it is unique and vary human to human naturally. In many applications like attendance and access, it is mostly used.
2. Facial Recognition: Many algorithms are developed to extract and match facial features like the shape of eyes, nose chin, and lips which are in appearance, and used for the authentication process of someone.

3. Iris: Depending upon the size and color of the Iris of the pupil, the pattern is used to verify someone in the process of authentication.
4. Palm: The palm is being used to control the system for example while capturing pictures in mobile phones and various smart applications the hand is identified and its gesture is used to control.
5. Geometry of Hand: Based on the size and structure of the hand, it is used in identity-based systems, for safety.
6. Signature: In day-to-day life, this factor is used in digital signature certificates issued by various service-providing companies for example the State and Central government-issued certificates, and trusted third parties.
7. Keystroke: For someone, his learning and writing patterns his keystrokes are used to validate identity.
8. Voice: The voice recognition mechanism is widely used as vocal words can be used as a factor in security.
9. Gait: The gait is the walking pattern, used in wearable IoT devices, to count the distance between feet, and also to show directions.

**Application:**

The Internet of Things is growing with several connected devices, and due to its versatile area of application, it has been implemented with success. It has given many benefits to society to improve productivity and lifestyles. Biometrics in IoT has been used with its full potential in domains like E-Health where the health of patients is monitored, also used for security in personal devices like mobile phones, smart homes, smart lockers, and social in societies of residence for example Smart Society, and it is widely adopted in industry places and logistic too. Nowadays due to the different hand gestures and signs it is used in identifying the sign language used in AI-controlled systems. Passports is major applications of Biometrics, used in authentication.

**Challenges & Future Direction:**

Biometric holds sensitive data, which can be misused, accuracy in matching the biometrics factors while authentication needs to more so as to be effective. Privacy of the collected data of biometrics needs to be ensured.

The biometric factors discussed in the papers are widely used in the process of authentication and security purposes. The number of devices using biometrics needs to be handled with care and proper security of the storage of devices needs to be done, as there are chances of data theft which can lead to huge security breaches. The market of the IoT is also growing, the biometrics and its implementation need to be done to increase security rather than to cut the cost of manufacturing. It is been observed number of attacks is growing and the data can be manipulated and changed as well by the hacker.

**Conclusion:**

The study of different types of biometric methods was examined. For the security of various devices and access these biometric factors are used and are mentioned. Future direction to the new researcher is also given to help them in research and contribute to the secure lifestyle of humans. Improving security in IoT also the user experience need to be done while achieving the goal of security.

**References:**

1. Mohammad S. Obaidat, Soumya P. Rana, Tanmoy Maitra, Debasis Giri, Subrata Datta, "Biometric Security and Internet of Things (IoT)", Chapter, 2018.
2. Wencheng Yang, Song Wang, Nor Masri Sahri, Nickson M. Karie, Mohiuddin Ahmed and Craig Valli, "Biometrics for Internet-of-Things: A Review", Sensor, 2021.
3. Chi-Wei lien and Sudip Vhaduri, "Challenges and Opportunities of User Authentication in the Age of IoT: A Survey", ACM, 2023.
4. Jyotsna Nalawade, "Finger Biometric for Internet of Things", International Journal of Advanced Research in Science, Communication and Technology, 2022.



## Automatic Sparse Representation for Classification of Echocardiographically Detected Intracardiac Masses

Dr. A.A. Tayade<sup>1</sup> and Prof. P.M. Ingle<sup>2</sup>

<sup>1</sup>Department of Computer Science, G.S. College Khamgaon, Dist. Buldhana

<sup>2</sup>Department of Computer Science, SSSKR Innani College Karanja Lad, Dist. Washim  
arvindtayade40@gmail.com<sup>1</sup>, inglepratik1@gmail.com<sup>2</sup>

### Abstract:

Identification of intracardiac masses in echocardiograms is one crucial duty in cardiac illness diagnosis. To increase diagnostic accuracy, a novel completely automated classification system based on the sparse representation is presented to identify intracardiac tumor and thrombi in echocardiography. First, a region of interest is chopped to determine the mass area. Then, a novel worldwide denoising procedure is utilized to eliminate the speckle and retain the anatomical structure. Subsequently, the contour of the mass and its linked atrial wall are represented using the K-singular value decomposition and a modified active contour model. Finally, the motion, the boundary as well as the texture data are processed by a sparse representation classifier to differentiate two masses. Ninety-seven clinical echocardiography sequences are collected to test the efficacy.

**Keywords:** *Echocardiographically, Intracardiac, Sparse Representation*

### I. Introduction

In recent years, there has been a growing interest in sparse representation. The sparse concept originated from the transform-domain methods, which assumed that true signals could be sparsely represented by a linear combination of few basis elements in the transform domain. Instead of using fixed and orthogonal transforms, images can be described by sparse linear combinations of an over complete dictionary.

Applications of the sparse representation include denoising, compression, regularization in inverse problems and classification. The K-Singular Value Decomposition (K-SVD) is one of the typical methods in the sparse representation, which utilizes over complete dictionaries obtained from a preliminary training procedure. Identification of intracardiac masses in echocardiograms is one important task in cardiac disease diagnosis. To improve diagnosis accuracy, a novel fully automatic classification method based on the sparse representation is proposed to distinguish intracardiac tumor and thrombi in echocardiography. Intra cardiac masses are abnormal structures found within or adjacent to the heart. These structures lead to severe cardiovascular disorders and require careful diagnosis for prompt resection and treatment. There are two main types of intra cardiac masses namely: tumor and thrombi. Tumor is the swelling of a part of body caused by abnormal growth of tissues which exhibits mobility and thrombi is a blood clot (solid mass of platelets).

Tumor is again divided into two: primary tumor and secondary tumor. Primary cardiac tumors are rare entities. They originate from the heart and can occur in any tissues of it. They can be cancerous or non-cancerous. Approximately 75% of them in adults are benign, with the majority composed of myxomas. It has an irregular shape and a gelatinous consistency. Secondary tumors are more common than primary. They do not originate in the heart. They move to the heart after developing in another area of the body. Cardiac tumours may not cause symptoms or may produce a severe cardiac dysfunction like sudden heart failure or sudden drop in blood pressure due to bleeding in pericardium. They cause obstruction to the left ventricular filling. Patients are present with the embolization, intra cardiac obstruction and constitutional sighs. Because of the high risk of embolization and sudden death, the tumours

need prompt resection. Intra cardiac thrombi are common findings in patients with ischemic stroke. This may lead to atrial fibrillation, enlarged atrial chamber and low cardiac outputs. Most patients with thrombi are treated with heparin and thrombolysis. For the non-invasive and low cost nature, echocardiography is widely used in diagnosis of intra cardiac masses. Echocardiography uses standard two-dimensional, three-dimensional, and Doppler ultrasound to create images of the heart. Echocardiography, also called an echo test or heart ultrasound, is a test that takes moving pictures of the heart with sound waves. You don't have to stay in the hospital.

The patient's heart movements can be seen on a video screen. The echocardiographic identifications of intra cardiac masses have great impacts on the medical doctors decision, since different diseases are related with diverse therapy options. In general, the echocardiogram sequence shows that most intra cardiac tumors have a narrow stalk and a broad base. The surface may be friable or villous. The internal echoes are heterogeneous. The tumors show continuity with the atrial wall, with a high degree of mobility. The echocardiographic appearances of the thrombi are motionless, dense, ovoid, and echo reflecting.

Although intra cardiac tumors and thrombi are different in pathology, they behave similarly in echocardiography. Often, they are misinterpreted. In most hospitals, echocardiographic identifications are carried out by cardiologists manually. The diagnosis is time-consuming. Recognition depends on the image quality and techniques, as well as the cardiologists experience. Hence, the demand for an automatic classification is increasing, which is potential to improve the diagnostic accuracy and to guide which case should be recommended for a surgery. The ultrasound image analysis has been successfully employed in the computer-aided diagnosis for cardiovascular disease, such as revealing valuable ultrasound features in early stroke prediction, designing fuzzy rule-based decision support system in the diagnosis of coronary artery, and applying adaptive block matching methodologies in carotid artery wall and plaque dynamics. Nevertheless, it is still challenging in intra cardiac masses identification due to the similar echocardiographic appearance of two masses and the suboptimal image quality including large amount of speckle noise, signal drop-out, artifacts, and missing contours. So a novel method is required to classify the intra cardiac tumor and thrombi in echocardiograms

## II. Literature Survey

**Yi Guo et al 2015** say that Identification of intracardiac masses in echocardiograms is one important task in cardiac disease diagnosis. To improve diagnosis accuracy, a novel fully automatic classification method based on the sparse representation is proposed to distinguish intracardiac tumor and thrombi in echocardiography. First, a region of interest is cropped to define the mass area. Then, a unique globally denoising method is employed to remove the speckle and preserve the anatomical structure. Subsequently, the contour of the mass and its connected atrial wall are described by the K-singular value decomposition and a modified active contour model. Finally, the motion, the boundary as well as the texture features are processed by a sparse representation classifier to distinguish two masses. Ninety-seven clinical echocardiogram sequences are collected to assess the effectiveness. Compared with other state-of-the-art classifiers, our proposed method demonstrates the best performance by achieving an accuracy of 96.91%, a sensitivity of 100%, and a specificity of 93.02%. It explicates that our method is capable of classifying intracardiac tumors and thrombi in echocardiography, potentially to assist the cardiologists in the clinical practice.

**Parisa Gifani al 2016** A challenging issue for echocardiographic image interpretation is the accurate analysis of small transient motions of myocardium and valves during real-time visualization. A higher frame rate video may reduce this difficulty, and temporal super resolution (TSR) is useful for illustrating the fast-moving structures. In this paper, we introduce a novel framework that optimizes TSR enhancement of echocardiographic images by utilizing

temporal information and sparse representation. The goal of this method is to increase the frame rate of echocardiographic videos, and therefore enable more accurate analyses of moving structures. For the proposed method, we first derived temporal information by extracting intensity variation time curves (IVTCs) assessed for each pixel. We then designed both low-resolution and high-resolution over complete dictionaries based on prior knowledge of the temporal signals and a set of prespecified known functions. The IVTCs can then be described as linear combinations of a few prototype atoms in the low-resolution dictionary. We used the Bayesian compressive sensing (BCS) sparse recovery algorithm to find the sparse coefficients of the signals. We extracted the sparse coefficients and the corresponding active atoms in the low-resolution dictionary to construct new sparse coefficients corresponding to the high-resolution dictionary. Using the estimated atoms and the high-resolution dictionary, a new IVTC with more samples was constructed. Finally, by placing the new IVTC signals in the original IVTC positions, we were able to reconstruct the original echocardiography video with more frames. The proposed method does not require training of low-resolution and high-resolution dictionaries, nor does it require motion estimation; it does not blur fast-moving objects, and does not have blocking artifacts

**O. Michailo vichet al 2002** Over the last few decades there were dramatic improvements in ultrasound imaging quality with the utilization of harmonic frequencies induced by both tissue and echo-contrast agents. The advantages of harmonic imaging cause rapid penetration of this modality to diverse clinical uses, among which myocardial perfusion determination seems to be the most important application. In order to effectively employ the information, comprised in the higher harmonics of the received signals, this information should be properly extracted. A commonly used method of harmonics separation is linear filtering. One of its main shortcomings is the inverse relationship between the detectability of the contrast agent and the axial resolution. In this paper, a novel, nonlinear technique is proposed for separating the harmonic components, contained in the received radio-frequency images. It is demonstrated that the harmonic separation can be efficiently performed by means of convex optimization. It performs the separation without affecting the image resolution. The procedure is based on the concepts of sparse signal representation in over complete signal bases. A special type of the sparse signal representation, that is especially suitable for the problem at hand, is explicitly described. The ability of the novel technique to acquire "un-masked," second (or higher) harmonic images is demonstrated in series of computer and phantom experiments.

### **III. Methodology**

In this method, a novel method is proposed to classify the intra cardiac tumor and thrombi in echocardiograms. Different from other approaches, the contribution of this method is incorporating the kernel collaborative region based classification into the algorithm. It involves the frame decomposition, automatic region of interests (ROIs) selection, globally despeckling, intracardiac mass segmentation, feature extraction, and classification. The video means multiple frames. The captured video is converted into frames using MATLAB codes. The cardiologists acquire echocardiogram sequences when diagnosing the disease. To segment the intracardiac mass and evaluate its movement, the echocardiographic sequences are divided into consecutive frames beforehand. The typical duration of an echocardiogram sequence is about 3–4 s. The frame rate is 39 frames per second. Each decomposed frame is  $480 \times 640$  pixels. Besides the scanned region, an echocardiogram depicts texts and labels, containing information about the patient and scanning transducer. Compared with moving heart in two successive frames, these texts and labels are static. After subtraction of two successive frames, the static information are all removed, while the sector scanned region containing moving heart is remained. Then, the profile of the sector scanned region is detected and a rectangle covering the sector is identified. Finally, the original image is cropped to keep the scanned region for further analysis

#### IV. Proposed Method

In this method, a novel method is proposed to classify the intra cardiac tumor and thrombi in echocardiograms. Different from other approaches, the contribution of this method is incorporating the kernel collaborative region based classification into the algorithm. It involves the frame decomposition, automatic region of interests (ROIs) selection, globally despeckling, intracardiac mass segmentation, feature extraction, and classification. The uniform sub windows in a coarse position, which, in turn, are searched to get a fine position with half size of sub windows. The iteration ends when all remaining sub windows share the same intensity distribution. In a short axis view echocardiogram, the chamber usually lies near the cardiac center. So in the fine position, the Euclidian distance between each sub window and the cardiac center is computed to trim off far-away windows and obtain final chamber location

#### V. Block Diagram

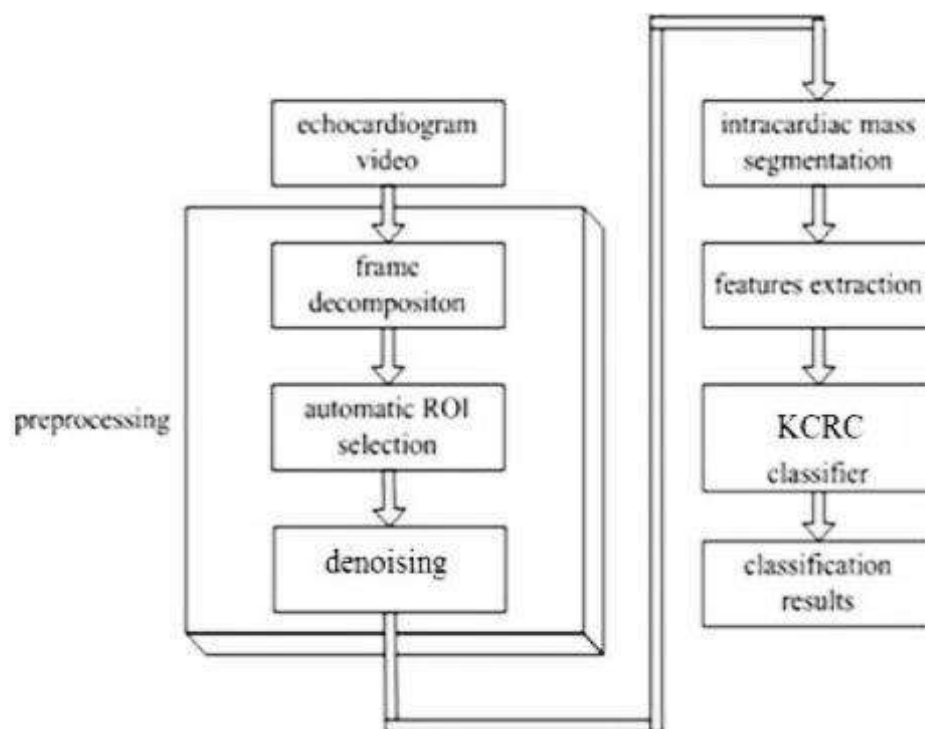


Fig. 1 Block Diagram of Workflow for the proposed classification method

#### VI. Conclusion

In this paper, a new method is proposed for the classification of intracardiac tumor and thrombi in the echocardiograms. The mass area in ROI is automatic defined by a coarse-to-fine strategy.. The denoising process yields better noise attenuation and edge enhancement, without destroying the important cardiac structures. The SLIC method is applied to segment the mass. Our detected contours closely approximate the manually traced ones. Nine features, including the cardiologist's original selected features and new better accuracy and simple implementation make the proposed method beneficial to help the cardiologists make a diagnosis before the surgery, providing a realistic performance benchmark for further research efforts.

---

**References**

1. X. Verbeek, J. Willigers, P. Brands, L. Ledoux and A. Hoeks, "Measurement of the contrast agent intrinsic and native harmonic response with single transducer pulse waved ultrasound system", *Ann. Biomed. Eng.*, vol. 27, pp. 670-681, 1999.
2. P. Frinking, A. Bouakaz, J. Kirkhorn, F. Ten Gate and N. de Jong, "Ultrasound contrast imaging: Current and new potential methods", *Ultrasound Med. Biol.*, vol. 26, no. 6, pp. 965-975, 2000.
3. D. H. Simpson, C. T. Chin and P. Burns, "Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 46, pp. 372-382, Mar. 1999.
4. J. Kirkhorn, P. Frinking, N. de Jong and H. Torp, "Three-stage approach to ultrasound contrast detection", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 48, pp. 1013-1022, July 2001.
5. K. E. Morgan, J. S. Allen, P. A. Dayton, J. E. Chomas, A. L. Klibanov and K. W. Ferrara, "Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 47, pp. 1494-1509, Nov. 2000.
6. P. Frinking, E. I. Cespedes, J. Kirkhorn and H. Torp, "A new ultrasound contrast imaging approach based on the combination of multiple imaging pulses and a separate release burst", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 48, pp. 643-651, May 2001.
7. Y. Li and J. A. Zagzebski, "Computer model for harmonic ultrasound imaging", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 47, pp. 1259-1272, Sept. 2000.
8. N. de Jong, P. Frinking, A. Bouakaz and F. Ten Gate, "Detection procedures of ultrasound contrast agents", *Ultrasonics*, vol. 38, pp. 87-92, 2000.
9. T. Christopher, "Finite amplitude distortion-based inhomogeneous pulse echo ultrasound imaging", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 44, pp. 125-139, Jan. 1997.
10. T. Christopher, "Experimental investigation of finite amplitude distortion-based second harmonic pulse echo ultrasound imaging", *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 45, pp. 158-162, Jan. 1998.
11. S. Chen and D. Donoho, "Atomic decomposition by basis pursuit", *SPIE Int. Conf. Wavelets*, 1995-July.
12. M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary", *Neural Computation*, vol. 13, no. 4, pp. 863-882, April 2001.

## EEG Signal Processing for Fetus and Mother Using MATLAB using Image Processing

**Dr. N.D. Jambhekar<sup>1</sup> and Dr. R.K. Nawasalkar<sup>2</sup>**

<sup>1</sup>Department of Computer Science, G.S. Gawande College, Umardhed, Dist. Yavatmal

<sup>2</sup>Department of Computer Science, G.S. Tompe College, Chandurbazar Dist. Amravati  
jambhekarnd@gmail.com<sup>1</sup>, ram\_nav78@rediffmail.com<sup>2</sup>

### Abstract :

During pregnancy, it is critical to diagnose the mother's and child's heartbeats, and fetal electrocardiogram (FECG) extraction is the method utilised to do so. During pregnancy and labour, the signal carries accurate information that might assist doctors. Independent component analysis has been used to build an easy-to-use technique. An efficient approach has been proposed using the ICA. For FECG extraction, the technique employs PCA and ICA. An algorithm written in MATLAB was used to implement the FECG extraction approach. The FECG signal that was recovered is noise-free. Post processing was utilised to detect the QRS complex, which used an adaptive noise filtering method to count the R-R peaks. The detection method can count the heart rate of the FECG signal, as shown in the end result. This project creates a complete FECG extraction model utilising effective algorithms and adaptive filters, and then produces the FECG data.

**Keywords:** ECG (Electrocardiography), FECG (Fetal Electrocardiogram), ICA (Independent Component Analysis), PCA (Principal component analysis)

### I INTRODUCTION

Every year, one in every one hundred kids is born with a heart defect. This can be caused by a genetic syndrome, an inherited condition, or environmental causes such as drug abuse. In any event, regular monitoring of the baby's heart is required prior to birth. As a result, Fetal ECG (FECG) signals are required to monitor the baby's heart status, so that any anomalies discovered can be treated clinically by the concerned specialists. Fetal ECG monitoring is a common method for detecting and diagnosing fetal abnormalities. By diagnosing the fetal ECG signal during the prenatal stage, the clinician can readily prepare himself or herself for any fetal anomalies. It's the simplest and least invasive way to diagnose a variety of cardiac conditions. The Fetal ECG (FECG) represents the heart's varied electrical activities and so provides useful information about its physiological state. The FECG signal can be easily collected from a pregnant woman's abdomen, whereas the maternal electrocardiogram (MECG) signal can be taken from her chest. The addition of the MECG signal to the FECG signal is usually a source of irritation. The FECG signal generated by putting electrodes on the maternal belly provides minute details about the fetal status that are particularly important during diagnosis. The maternal ECG (MECG) signal and electromyogram (EMG) signal are contaminated by various noise and skin impedance, whereas the FECG signal is contaminated by various noise and skin impedance. The electrocardiogram (ECG) is a method of describing the electrical activity of the heart. Three basic types of waves make up ECG signals. The heart rate of the abdominal ECG (AECG) signal is determined by the peaks value of the QRS complex. As a result, it is critical for doctors to diagnose heart problems before they cause harm to the fetus or the mother. When the ECG signals from the abdomen leads are combined, a composite signal is created. Adaptive filtering, wavelet Transform, Independent component Analysis, Principle Component Analysis, Fatal ECG Extraction from Maternal Abdominal ECG Using Neural Network, Fetal ECG Extraction for Fetal Monitoring Using SWRLS Adaptive Filter and Extraction of Fetal ECG from Maternal ECG using Least Mean Square Algorithm are some of

the methods for extracting FECG from AECG. Every year, one in every one hundred kids is born with a heart defect. This can be caused by a genetic syndrome, an inherited condition, or environmental causes such as drug abuse. In any event, regular monitoring of the baby's heart is required prior to birth. As a result, Fetal ECG (FECG) signals are required to monitor the baby's heart status, so that any anomalies discovered can be treated clinically by the concerned specialists. Fetal ECG monitoring is a common method for detecting and diagnosing fetal abnormalities. By diagnosing the fetal ECG signal during the prenatal stage, the clinician can readily prepare himself or herself for any fetal anomalies. It's the simplest and least invasive way to diagnose a variety of cardiac conditions. The Fetal ECG (FECG) represents the heart's varied electrical activities and so provides useful information about its physiological state. The FECG signal can be easily collected from a pregnant woman's abdomen, whereas the maternal electrocardiogram (MECG) signal can be taken from her chest. The addition of the MECG signal to the FECG signal is usually a source of irritation. The FECG signal generated by putting electrodes on the maternal belly provides minute details about the fetal status that are particularly important during diagnosis. The maternal ECGb(MECG) signal and electromyogram (EMG) signal are contaminated by various noise and skin impedance, whereas the FECG signal is contaminated by various noise and skin impedance. The electrocardiogram (ECG) is a method of describing the electrical activity of the heart. Three basic types of waves make up ECG signals. The heart rate of the abdominal ECG (AECG) signal is determined by the peaks value of the QRS complex. As a result, it is critical for doctors to diagnose heart problems before they cause harm to the fetus or the mother. When the ECG signals from the abdomen leads are combined, a composite signal is created.

## **II BACKGROUND**

Manisha Dodatale et al 2021 says that During pregnancy, it is critical to diagnose the mother's and child's heartbeats, and fetal electrocardiogram (FECG) extraction is the method utilised to do so. During pregnancy and labour, the signal carries accurate information that might assist doctors. Independent component analysis has been used to build an easy-to-use technique. An efficient approach has been proposed using the ICA. For FECG extraction, the technique employs PCA and ICA. An algorithm written in MATLAB was used to implement the FECG extraction approach. The FECG signal that was recovered is noise-free. Post processing was utilised to detect the QRS complex, which used an adaptive noise filtering method to count the R-R peaks. The detection method can count the heart rate of the FECG signal, as shown in the end result. This project creates a complete FECG extraction model utilising effective algorithms and adaptive filters, and then produces the FECG data

Esha Ahuja et al 2016 The extraction of Fetal Electrocardiogram (FECG) is vital to know the well being of the fetus and useful for doctors to decide the mode of delivery and period. The FECG contains activity of electrical depolarization and repolarization of fetal heart. In this paper, a simple algorithm, Independent Component Analysis (ICA), is used to extract FECG from Abdominal Electrocardiogram (AECG) of mother. The database used is non-invasive fetal electrocardiogram and direct fetal electrocardiogram, taken from physionet.org. ICA comes under the classification of Blind Source Separation (BSS) method. ICA is basically a filtering solution which gives the signal from an unknown source. In this problem, the signal from an unknown source is Fetal ECG. It is to be derived from the pure maternal ECG that is thorax signal and abdominal ECG

Md. Kafiul Islam et al 2020 EEG recordings are usually affected by various artifact types come from non-neural sources and make it difficult for accurate signal classification in the later stage. Thus reliably detecting and removing artifacts from EEG by an automated signal processing algorithm is an active research area. In this paper we have developed a wavelet based artifact removal algorithm from EEG data that selects the best (optimal) threshold parameters, and hence consequently provides the best performance of artifact removal. In the proposed

algorithm we choose to sweep both the wavelet filter parameter and threshold parameters until the best accuracy and/or least distortion is achieved by making a decision based on a reference dataset. The criteria for optimized selection are based on the metrics that quantify both amount of artifact removal and amount of distortion in the signal in both time and frequency domain. The algorithm is tested on synthesized EEG data that include different artifact templates and thus quantifies the performance based on several time and frequency domain measures. The achieved results prove that by selecting the optimum mother wavelet and parameter values adaptively would give the best performance both with regard to amount of artifact removal and least signal distortion compared with selecting any predefined mother wavelet and/or constant threshold parameter. This research would help the EEG signal analysis community a platform to work further in future on such problem to be able to properly select the wavelet parameters.

### **III Adaptive Filtering Based FECG Extraction**

An adaptive filter is one that self-adjusts its transfer function in response to an error signal that drives an optimization process. For the separation of fetal and maternal signals, various adaptive filters have been applied. These approaches extract the deadly QRS waves by training an adaptive or matching filter with one or more reference maternal signals[1][2]. For MECG cancellation and FECG extraction, the kalman filter, a broad form of adaptive filter, requires simply an arbitrary MECG as a reference. The temporal dynamics of AECG signals were synthesized using a collection of state-space equations and a Bayesian filter, which was employed for ECG de noising. However, when the maternal and deadly components are completely overlapped in time, the filter is unable to distinguish between them. When the waves of mixed signals fully overlap in time, it is said to be wholly overlapped. This filter makes it difficult to filter out the required ECG. A superior strategy was proposed in Buses multistage adaptive filtering for FECG extraction, where MECG cancellation was conducted using thoracic ECG as a reference signal, and de noising methods were used to improve the quality of the resultant signal. Normally, an adaptive filter requires two input signals (AECG signal and Thoracic ECG), but in this case, the thoracic ECG has been scaled and squared. Adaptive filters were well calibrated to conduct the extraction once scaling factors were chosen. The advantage of this technique is that the input thoracic signal does not have to be original, as it was collected from a pregnant lady whose AECG was also provided as primary input; alternatively, a signal that is very similar can be used. This self-adjusting filter uses three alternative methods to boost the SNR ratio by altering filter coefficients: LMS, RLS, and NLMS. The FECG was extracted using a linear adaptive filter[3]. The FECG was extracted using abdominal ECG as primary inputs and thoracic ECG collected from the mother's chest as reference inputs. Despite the fact that the offered method gives a solution, it fails to extract when maternal and lethal signals coincide. As a result, it is not appropriate for clinical use.

### **IV METHODOLOGY**

The subjective signal and the wavelet function are convolutioned in the Wavelet Transform. The Wavelet Transform divides a signal into two parts, the detail signal and the approximation signal. The detail signal is found in the upper half of the frequency component, while the approximation signal is found in the lower half. In the discrete wavelet domain, multi-resolution analysis is thus possible. A significant number of well-known wavelet families and functions are available for a wide range of applications. Bi-orthogonal, Coiflet, Harr, Symmlet, and db (Daubechies) wavelet are some of the wavelet families. The wavelet function is utilized depending on the application. These wavelet families have been used in a variety of research projects. It is not possible to select a certain wavelet. To obtain the wavelet analysis, we use the MATLAB application. The wavelet toolkit in MATLAB is quite extensive. We employ the db (Daubechies) wavelet in the algorithm because it resembles the waveform of a human heart beat outcome of the daubechies is excellent. In this research, we employ the daubechies wavelet transform to construct an algorithm in MATLAB that decomposes the



signal into approximation and detailed coefficients. The Wavelet Transform is based on the convolution of the subjective signal and the wavelet function. An automated approach is used to extract the Fetal Electrocardiogram (FECG) from the Abdominal Electrocardiogram. In [23], a (AECG) recording was described. The information was gathered in a non-invasive manner. Data is collected using external electrodes inserted on the abdomen. An AgAgCl transducer is used to boost SNR, and electrode positions are altered. To capture signals, electrodes were placed on the mother's abdominal wall. The.edf format of the abdominal electrocardiogram (AECG) data employed in this technique was converted to a MATLAB readable format. There are three main procedures that are considered. Pre-processing, FECG extraction and post-processing are all steps in the process. First read the recorded mothers Abdominal ECG signal in MATLAB and then use PCA to eliminate undesirable noise from the AECG signal. After preprocessing, the FECG signal was extracted using the ICA approach. Then R-Peak is discovered, and Fetal Heart Rate Calculation is performed.

## V.BLOCK DIAGRAM

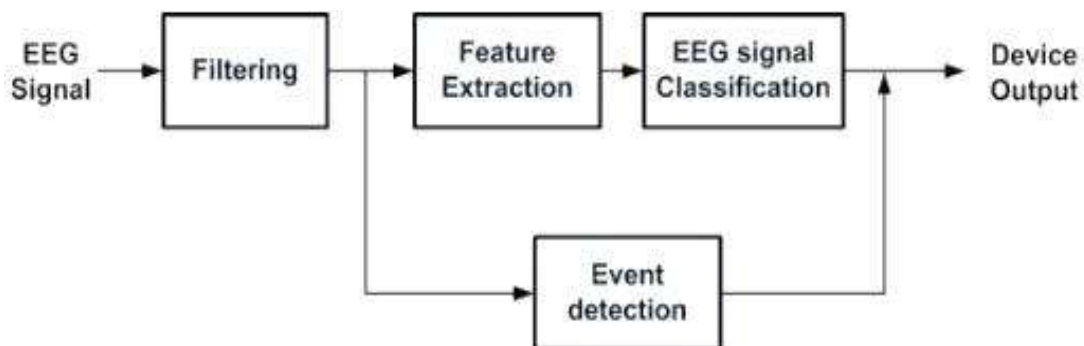


Fig 1: Block Diagram of the overall system

## VI INTERPRETATION

- EEG Signal are functionally fast, relatively cheap and safe way of checking the functioning of different areas of brain.
- High precision time measurements
- Today's EEG technology can accurately detect brain activity at a resolution of a single millisecond.
- EEG electrodes are simply stuck onto the scalp. It is therefore a non-invasive procedure.
- EEG equipment is relatively inexpensive compared with other devices and is simple to operate. An automated approach is used to extract the Fetal Electrocardiogram (FECG) from the Abdominal Electrocardiogram. In [23], a (AECG) recording was described. The information was gathered in a non-invasive manner. Data is collected using external electrodes inserted on the abdomen. An AgAgCl transducer is used to boost SNR, and electrode positions are altered. To capture signals, electrodes were placed on the mother's abdominal wall. The.edf format of the abdominal electrocardiogram (AECG) data employed in this technique was converted to a MATLAB readable format. There are three main procedures that are considered. Pre-processing, FECG extraction and post-processing are all steps in the process. First read the recorded mothers Abdominal ECG signal in MATLAB and then use PCA to eliminate undesirable noise from the AECG signal. After preprocessing, the FECG signal was extracted using the ICA approach. Then R-Peak is discovered, and Fetal Heart Rate Calculation is performed.

## V CONCLUSION

The fatal electrocardiogram (FECG) has a humble beginning dating back to 1901, when the first expansion of research in the associated arena was severely limited. The identification of the waveform was greatly eased with the introduction of improved amplifiers and filters, yet waveform morphology surveillance was a difficult issue due to the presence of background noise after the contaminated signal was filtered. The signal-to-noise ratio of the original FECG was dramatically improved thanks to sophisticated processing and computer technologies, notwithstanding the signals' non-invasive acquisition. The document beautifully displays the evaluation of a variety of methodologies that have been widely used for the extraction of FECG up to this point.

## REFERENCES

1. X. Shen, J. Wu, Y. Zhang, Y. Li, and Y. Ma, "Towards an evaluation model of online learning behavior and learning effectiveness for moocap learners," Distance Education in China, vol. 7, 2022..
2. A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, "Neonatal seizure detection using deep convolutional neural networks", Int. J. Neural Syst., pp. 1850011, 2022
3. H. Abbasi, A. Gunn, L. Bennet and C. Unsworth, "Deep convolutional neural network and reverse biorthogonal wavelet scalograms for automatic identification of high frequency micro-scale spike transients in the post-hypoxic-ischemic EEG", EMBC, vol. 20, 2020
4. O'Shea, G. Lightbody, G. Boylan and A. Temko, "Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture", Neural Networks, vol. 123, pp. 12-25, 2021..
5. H. Abbasi, A. Gunn, C. Unsworth and L. Bennet, "Wavelet spectral time-frequency training of deep convolutional neural networks for accurate identification of micro-scale sharp wave biomarkers in the post-hypoxic-ischemic EEG of preterm sheep", EMBC, vol. 20, 2020..
6. H. Abbasi and C. P. Unsworth, Applications of advanced signal processing and machine learning in the neonatal hypoxic-ischemic electroencephalogram, vol. 15, no. 2, pp. 222-231, 2020
7. B. Bateman, A. R. Jha, B. Johnston and I. Mathur, A New Interactive Approach to Understanding Supervised Learning Algorithms, Birmingham, U.K.:Packt Publishing, pp. 342-346, 2020.
8. A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, "Neonatal seizure detection using deep convolutional neural networks", Int. J. Neural Syst., pp. 1850011, 2018.
9. X. Shen, J. Wu, Y. Zhang, Y. Li, and Y. Ma, "Towards an evaluation model of online learning behavior and learning effectiveness for moocap learners," Distance Education in China, vol. 7, 2019..
10. H. Abbasi and C. P. Unsworth, Electroencephalogram studies of hypoxic-ischemic encephalopathy in fetal and neonatal animal models, vol. 15, no. 5, pp. 828-837, 2020

**Vidya Bharati Mahavidyalaya, Amravati**



Published by : SAI JYOTI PUBLICATION  
Behind Chawla Sadi Center, Tin-nal Chowk,  
Kasarpura, Itwari, Nagpur-440002  
Phone : 9764673503, 9923593503  
email : sajp10ng@gmail.com  
Website : www.saijyoti.in

