PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INFORMATION PROCESSING AND COMPUTING – 2022

(ICIPC – 2022)
27th – JUNE – 2022

Chairman

Prof. Dr. Vijayan Sugumaran,

Co-Chair and Treasurer

Prof. Dr. S. Jafar Ali Ibrahim,

Convener

Prof. Dr. R. Suneetha Rani

Royal Book Publishing

Book Title	PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INFORMATION PROCESSING AND COMPUTING – 2022
Book Size	6.69 * 9.84 Inch
Paper	Natural Shade
Publisher	 **Royal Book Publishing** **21/11, K M Nagar,** **Ayodhiyapatinam,** **Salem.** **Tamil Nadu-636103** **Cell:+91 7598141199** **Email:** contact@royalbookpublishing.com **Website:** www.royalbookpublishing.com

---

ISBN Assigned by Raja Ram Mohun Roy National Agency for ISBN,

New Delhi – 110066 (India)

ISBN: 9789391131814

President Message

Sri N.Nageswara Rao
B.E., MIE., President

President – SNES

As we face the end of the first decade of the new millennium, we must come together to share ideas and resources so we can plan and create a better future for the next generation of students, faculty, and researchers. In the tight, competitive global markets of the 21st century, the leading companies in various industries have embarked on massive reorganizations, mergers, partnerships, and cutting-edge collaborative projects with their like-minded peers—including their rivals—primarily to survive but ultimately to grow.

I hope this (ICIPC-2022) conference will reshape and create a better future for the next generation of students, faculty, and researchers in academia and industry.

My best wishes for this conference's grand success.

Secretary & Correspondent Message

Dr. N.S.Kalyan Chakravarthy
M.Tech., Ph.D.

Secretary & Correspondent
QIS Educational Institutions

The theme of the conference is "Recent Trends in Engineering and Technology". This theme has been chosen after careful deliberation, to reflect the fact that there have been fundamental changes to both the character and the dynamics of the global industrial need since the turn of the century, and that this may, in turn, affect the way the industry addresses the challenges that lie before it.

Today, technologies like AI, IoT, big data, 5G, autonomous robots, and blockchain are stand-alone solutions. It is not a small task to ensure a variety of IoT sensors can speak with a manufacturing execution system, which is, in turn, able to talk with a cloud-based data analytics package. That leaves producers with two choices: They can either find a vendor who packages all these capabilities together, though this may lock them into a single and often expensive proprietary system. Or, if they want to mix and match best-of-class applications, they must pay programmers to integrate devices and software, so data formats are compatible up and down the system.

I am very proud to invite global researchers to join hands with our QIS Educational Institutions through this conference to carry out research and make new discoveries that are beneficial to mankind in the years to come.

I am delighted to declare open this conference on Recent Trends in Engineering and Technology, and I hope this conference will share more knowledge with all. Also, I greet this conference will be a successful one.

Principal Message

Dr. Y. V. Hanumantha Rao, M.Tech., PhD
Principal

The sector of Mechanical Engineering is the primary consumer of Artificial Intelligence as a technology. It is more than any other industry; it is consumed the most in Mechanical designs or engineering works. Sections of Mechanical Engineering like Robotics, Automation, or sensor technology, use Artificial Intelligence as a technology. So, it is easy to say that Mechanical Engineering disseminates the application and use of AI in the ecosystem.

Artificial Intelligence isn't a long-listed dream anymore. More and more industries are taking advantage of it and providing some fantastic results that can help the human ecosystem. Yes, AI might affect an industry's HR by eradicating the least essential employee. Yes, it might kill the human element attached to the process. Yet, against all of it, it is helping our generation to imagine an entirely new world by impacting the different fields.

I hope this conference will be a good platform for all of you for advancing your knowledge.

Key Note Speaker – 1

Prof. Dr. Vijayan Sugumaran

Distinguished Professor, Management Information Systems

Chair, Department of Decision and Information Sciences

Co-Director, Center for Data Science and Big Data Analytics

School of Business Administration

Oakland University, Rochester, Michigan,

United States of America.

Key Note Speaker – 2

Dr. M. Sathya , B.E, M.Tech, MBA,.

Professor and Head,

Department of Information Sciences

AMC Engineering College,

Bangalore.

Key Note Speaker – 3

Dr. Gulfam Ahamad
Assistant Professor, Department of Computer Sciences,
Baba Ghulam Shah Badshah University, Rajouri,
Jammu and Kashmir-185234

Table of content

Enhancement of 3-D Pythocrypt using AES Technique by generating the key using Chitra Kavya

Sumanth N
Dept. of ISE
Jyothy Institute of Technology
Bengaluru, India
nsumanthbhat@gmail.com

Dr. Padmashree Anand
Ancient Science and Technology
Center for Incubation, Innovation, Research and consultancy(CIIRC)
Bengaluru, India
padmashree.a@ciirc.jyothyit.ac.in

Shrinidhi Holla
Dept. od ISE
Jyothy Institute of Technology
Bengaluru, India
hollashrinidhi99@gmail.com

Prof. Rakshatha S
Asst. Professor, Dept. of ISE
Jyothy Institute of Technology
Bengaluru, India
rakshatha.s@jyothyit.ac.in

Prajwal VS
Dept. of ISE
Jyothy Institute of Technology
Bengaluru, India
prajwal.vs@gmail.com

Dr. Harshvardhan Tiwari
HOD, Dept. of ISE
Jyothy Institute of Technology
Bengaluru, India
hod.ise@jyothyit.ac.in

Abstract - The research work describes an attempt to enhance the 3-D Pythocrypt algorithm using the AES Symmetric encryption technique by generating the key using an ancient approach of Chitra kavya. 3-D Pythocrypt is a new technique where we use the properties of 3-D geometric shapes to encrypt and decrypt plain text. AES encryption is a standard approach for encrypting and decrypting the message using a single key for both approaches. The ideology of this research is to first encrypt the plain text using the 3-D Pythocrypt technique and the cipher text obtained is fed as an input to the AES encryption technique along with the key generated using the Chitra kavya to carry out the encryption in two phases.

Keywords - 3-D Pythocrypt Algorithm, AES Technique, Symmetric Encryption, Decryption, Key generation, Chitra kavya, Cipher text, plain text.

I. INTRODUCTION:

The present web world is overwhelmed with a gigantic measure of information that is coming about information spillage and robbery of data. Bunches of difficulties emerge while moving or getting different sorts of encrypted information, messages, or data particularly utilizing public organizations. Digital wrongdoing is one of the greatest blemishes in this completely associated internetworking world. We present the strategy of enhancing 3-D Pythocrypt using the AES Symmetric encryption technique. 3-D Pythocrypt is a new technique, where the properties like area, volume, perimeter, etc. of specific 3-D geometric shapes are used to encrypt and decrypt the message which needs to be transmitted to attain confidentiality, AES Encryption technique is a traditional technique that uses the same key for both encryption and decryption. In this research, the key is generated using a different approach i.e., the Chitra kavya technique.

A. 3-D Pythocrypt:

There is a huge number of cryptographic systems which can be used to encrypt the message and transmit the cipher text that is generated so that the confidential information is made readable only by the intended user. Traditionally cryptographic systems are classified into two categories,

Symmetric, and Asymmetric cryptographic systems. In Symmetric cryptographic systems, the same key is used for both encryption and decryption process, some of the popular symmetric techniques like the Caesar cipher model, play fair cipher model, and DES and AES Encryption model. The asymmetric cryptographic system uses different keys for encryption and decryption processes called public keys and private keys, the most popular and widely used asymmetric algorithm is RSA.

3-D Pythocrypt is a new and young technique that is used to encrypt and decrypt a given message using a known 3-D geometric shape (Jois et al., 2015). This methodology uses the geometric attributes of shapes such as Area, Volume, Perimeter, etc., to generate an algorithm using which the plain text can be encoded into non-decipherable text. This new approach is unique compared to all other existing algorithms as it does not need a key for encryption but generates one for decryption [1].

B. AES Symmetric Encryption Technique:

Advanced Encryption Standard is a symmetric encryption technique i.e., it uses the same key for both encryption and decryption process. The key needs to be kept secret for the algorithm to be secure. AES algorithm can be implemented using different key lengths i.e., AES-128 which uses a key of length 128 bits, AES-192 where the length of the key is 192 bits, and AES-256 where the key length is 256 bits (key length is 32 digits) [2]. It uses substitution and permutation techniques to encrypt and decrypt the plain text [3].

C. Chitra Kavya:

Sanskrit poetry encompasses a variety of forms and structures which include Mahākāvya, Laghukāvya, Khandakāvya, Campukāvya, Gītikāvya, Muktakas, Stotras, Biographical Poems, Citrakāvya, etc. Citrakāvya is a wonderful class of poetry based on the intricate and innovative play of vowels, consonants, words, and sounds. Citrakāvya has a figurative quality wherein the elements of the verses are picturesquely patterned into designs resembling objects (bandhas) or their movements (gati) such as flower, wheel, flag, drum, etc. Also, the term Citra connotes an image or picture, uniqueness (vicitra), or wonder. Chitrakāvya is interpreted as 'image poetry' or 'marvel poetry' which embraces all ingenious forms of poetic compositions.

Ākaracitra and Bandhacitra are techniques by employing which verses can be designed and woven into various patterns of objects, animals, birds, etc. There are more than 200 known varieties of Bandhas. These include 12 types of Nāga-bandha of single or multiple coiled or uncoiled snakes; 19 types of Āyudhas, weapons such as sword, knife, mace, and others; 16 types of Ābharaṇa-Chitra's resembling ornaments such as bangle, armlet, girdle, etc; and 38 types of miscellaneous formations - Anya-Ākāra-*Citra:* those resembling umbrella (chatrabandha), banner or flag (patākabandha), mace (gadābandha) bow (dhanurbandha; Figure 2) in addition to sun, moon, Meru, bed, swing, lamp, pestle, bell and so on. Fig. 1 is the parashu bandha, in which the letters of the verse is placed in a particular sequence. We can decrypt the verse (shloka) only when the traversing pattern is known.

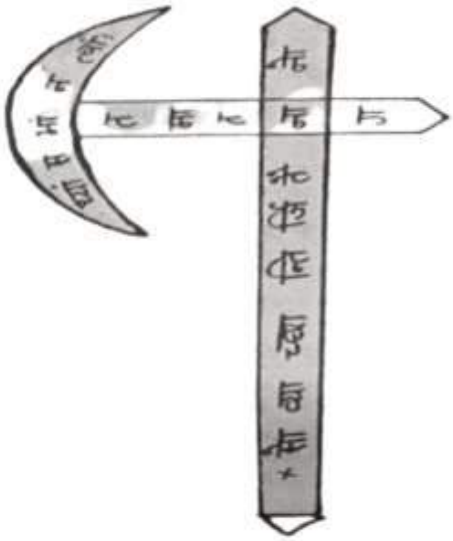


Fig. 1. Parashu Bandha

II. PROPOSED SYSTEM:

The system combines the two different techniques i.e., 3-D Pythocrypt and AES Symmetric algorithm to form a complex system that has two layers of encryption to provide better security against cryptanalysis. The plain text is

converted to cipher text using the young cryptographic technique called 3-D Pythocrypt algorithm. The cipher text obtained from this process is passed to the AES encryption technique along with the key generated by the Chitra kavya technique, the output of this AES technique is used for transmission or storing the data.

III. ALGORITHM:

A. Encryption Phase:

3-D Pythocrypt Encryption: The 3-D Pythocrypt algorithm is based on the properties of 3-D geometric shapes. This algorithm takes the properties like Area, Volume, etc. of the shapes. The message which consists of English alphabets and/or symbols is converted to its ASCII equivalent values which are in decimal format. These numbers are halved and considered as inputs to the formulae to calculate the ciphertext. In our research work, we have used different 3-D shapes to make the algorithm more secure. Whenever a user inputs a plain text, its corresponding ASCII value is generated. The entire ASCII value is halved. There are two variables in the formula namely 'a' and 'h'. The halved ASCII values are the inputs for these variables. For a particular plain text, a shape is chosen and the encryption operation is performed. The result is the cipher text obtained from the 3-D Pythocrypt encryption phase. Along with the cipher text, any one value of the variable 'a', 'a2', or 'h' can be used as the key. Table 1. Contains the different 3-D geometric shapes with their respective volumetric formula. Whenever the user inputs a new plain text, a particular shape and formula from Table 1. Is chosen and is used for the encryption process. Id column in Table 1. Contains the id's assigned to the shapes which will help us to communicate the shape used at encryption to the receiver.

TABLE 1. DIFFERENT 3-D SHAPES AND THEIR CORRESPONDING VOLUMETRIC FORMULA.

Id	Shape	Volumetric Formula
1	Octahedron	$v = (2 * a^2 * h)/3$
2	Hexagonal Prism	$v = (3 * \sqrt{3} * a^2 * h) / 3$
3	Pentagonal Prism	$v = (\sqrt{(5(5+2\sqrt{5}))} * a^2 * h) / 4$
4	Octagonal Prism	$v = 2 * (1 + \sqrt{2}) * a^2 * h$
5	Pentagonal Pyramid	$(5 * \tan(54^o) * h * a^2)/12$

AES Encryption Algorithm: The cipher text generated above along with the key needs to be transmitted to the receiver end. To enhance the security of the 3-D Pythocrypt algorithm, we encrypt the cipher text and the key using the AES symmetric encryption technique. However, the cipher text and the key needs to be combined using a delimiter. Here we use '.' As the delimiter hence, the actual cipher text from the 3-D Pythocrypt algorithm will be like [Output_of_algorithm.key] ex: 12345.5678. This together is passed as the input for the AES encryption algorithm.

Key Generation for AES Encryption: AES is a symmetric encryption technique which means it uses the same key for both encryption and decryption purposes. For generating the key for the AES encryption phase, we introduce a different technique called the Chitra kavya technique. As mentioned earlier, Chitra kavya contains different designs (bandha) where the verse (shloka) is encrypted in those designs. If we traverse those designs (bandha) in a particular order we can decrypt the verse(shloka/message). Here we construct a 21 x 21 matrix (dimension is not fixed, it can be altered) in which the elements are an orderly arrangement of numbers from 0 to 9 (as shown in Fig. 2.).

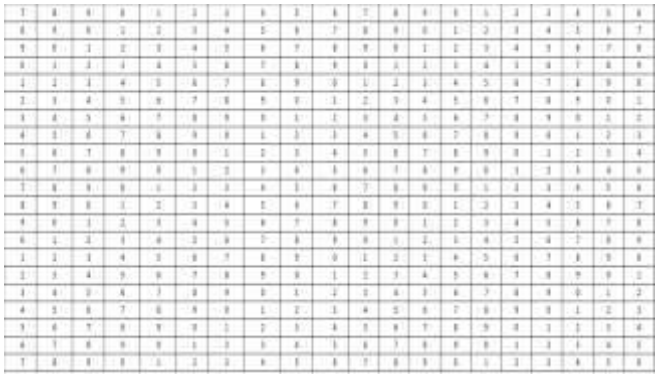


Fig. 2. A Sample 21 x 21 matrix

A dynamic seed point is obtained as the initial node and the matrix is traversed in a specific order such that the path forms the chosen design. In this research work, we have considered ***parashu bandha(design).*** We consider element 2 (5,13) (5th row and 13th column) as the seeding point. From this initial node, we traverse through the matrix in such a way that the path from the parashu bandha (Fig. 3.)

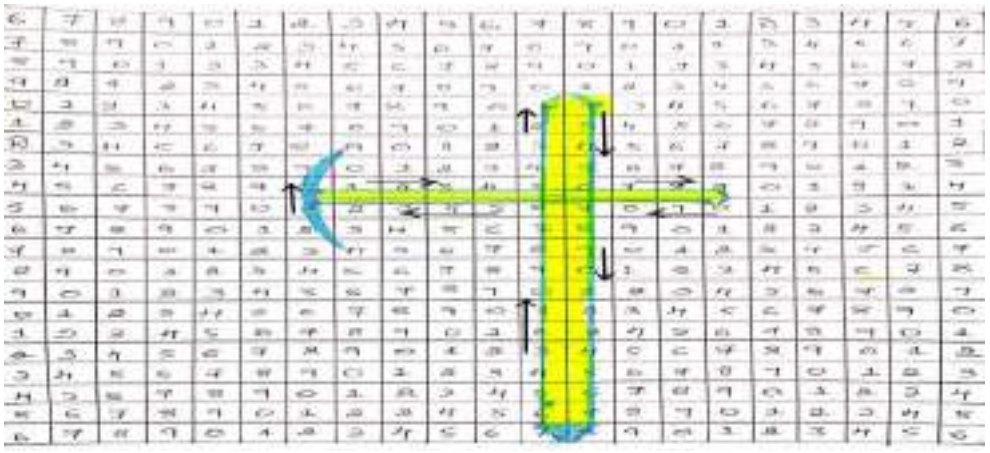


Fig. 3. Traversing the matrix to form parashu bandha(design).

The values encountered while traversing is taken as the key for the AES symmetric encryption algorithm. From fig. 3, we generate the key as 234567890987890123456787654321098765432124999012345432 1. But, for the AES encryption, we only need a 256-bit key. Hence, we consider only the first 32 digits of the above-generated value as the secret key for the encryption and decryption process.

The output of the AES encryption algorithm is transmitted to the receiver.

B. Decryption Phase:

On the receiver side, the received message(data) is first decrypted using the AES Symmetric decryption technique. The result of this is the cipher text obtained from the 3-D Pythocrypt algorithm which is decrypted to get the original plain text.

AES Decryption Algorithm: The encrypted cipher text of the 3-D Pythocrypt algorithm along with the seeding point used in generating the key is fed as an input to the AES Decryption algorithm. Using this seed point, the key is generated at the receiver side. Hence, the key is not transmitted from the sender to the receiver however, some knowledge about the key generation (matrix dimension, the design(bandha) used for encryption) should be known to the receiver. Using the key and the message received, the cipher text obtained from the 3-D Pythocrypt encryption technique is recovered.

3-D Pythocrypt Decryption: All the information that is required to decrypt the cipher text (geometric shape used, key) is embedded in the cipher text. The key is extracted from the cipher text using the delimiter. Here the delimiter is '.'. As mentioned earlier, any one value of variables 'a', 'a2', or 'h' can be used as the key. The decryption formula changes accordingly. If we use 'a' or 'a2' as the key then, we need to find 'h'. However, if we use 'h' as the key then, we need to find the value of 'a'. Table 2. Describes the different equations that need to be used to calculate the unknown value.

TABLE 2. 3-D PYTHOCRYPT DECRYPTION FORMULAE LIST.

Id	Shape	Decryption Formula		
		if key = a	If key = a^2	If key = h
1	Octahedron	$a^2 = (a * a)$ $h = (3 * v)/(2 * a^2)$	$h = (3 * v)/(2 * a^2)$	$a^2 = (3 * v)/(2 * h)$ $a = \sqrt{a^2}$
2	Hexagonal Prism	$a^2 = (a * a)$ $h = (2 * v)/(3 * \sqrt{3} * a^2)$	$h = (2 * v)/(3 * \sqrt{3} * a^2)$	$a^2 = (2 * v)/(3 * \sqrt{3} * h)$ $a = \sqrt{a^2}$
3	Pentagonal Prism	$a^2 = (a * a)$ $h = (4 * v)/\sqrt{(5(5+2\sqrt{5}))} * a^2$	$h = (4 * v)/\sqrt{(5(5+2\sqrt{5}))} * a^2$	$a^2 = (4 * v)/\sqrt{(5(5+2\sqrt{5}))} * h$ $a = \sqrt{a^2}$
4	Octagonal Prism	$a^2 = (a * a)$ $h = v/((2 * (1+\sqrt{2}) * a^2)$	$h = v/((2 * (1+\sqrt{2}) * a^2)$	$a^2 = v/((2 * (1+\sqrt{2}) * h)$ $a = \sqrt{a^2}$
5	Pentagonal Pyramid	$a^2 = (a * a)$ $h = (12 * v)/(5 * \tan(54^0) * a^2)$	$h = (12 * v)/(5 * \tan(54^0) * a^2)$	$a^2 = (12 * v)/(5 * \tan(54^0) * h)$ $a = \sqrt{a^2}$

Once 'a' and 'h' values are successfully retrieved, these two are combined to form the ASCII values of the original plain text. Eventually, the original plain text is extracted from its ASCII values.

IV. IMPLEMENTATION:

The design of the system contains two parts, the 3-D Pythocrypt and AES Symmetric algorithm. The message which needs to be encrypted is fed as an input to the 3-D Pythocrypt algorithm. The Pythocrypt algorithm converts this message to the ciphertext which is difficult to understand and also the key is generated. This will be in decimal format. The generated ciphertext (encrypted message**.** key) is fed as input to the AES encryption algorithm, which encrypts the cipher text by substitutions and permutations using the key generated using the Chitra kavya technique.

The decryption phase is the reverse of the encryption phase. The encrypted message from the AES encryption algorithm is passed to the AES decryption algorithm to retrieve the cipher text. This cipher text is passed to the 3-D Pythocrypt decryption algorithm to get the original plain text. Fig. 4 is the pictorial representation of the flow of data (plain text) from the sender to the receiver.

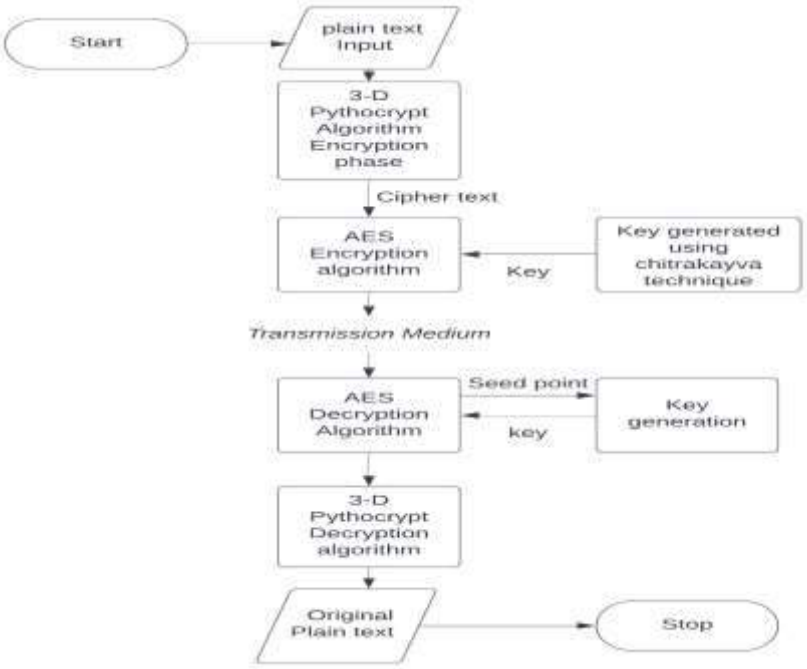


Fig. 4. A Flowchart that describes the entire process.

RESULT:

- The plain text can be a sequence of any character which has equivalent ASCII values.
- For example: Let the plain text be, "**Hello, this is a sample message**".
- This sample plain text is 31 characters long, and its equivalent ASCII value will be 721011081081114432116104105115321051153297321159710911210810132109101111511597103101.
- When this plain text is encrypted using the volume of octahedron geometric shape, the corresponding cipher text obtained from the 3-D Pythocrypt algorithm will be, 131113047447361507825287389109365958728990393326220064151394961889866415993395386488073700030234557263079170189735567059701207 3.32115971091121081013210910111515971031010

- Note that the dot (".") operator is used as the delimiter between cipher text and the key. Both cipher text and the key will be sent to the AES Encryption algorithm.
- From Fig.3.the key for AES encryption can be 234567890987890123456787654321098765432124999012345432 1,
  However, for the AES encryption 256-bit (32 digits) key is sufficient. Hence, we consider only the first 32 digits of the above-generated key (23456789098789012345678765432109).
- The cipher text and the 32-digit key are passed as the input to the AES Encryption algorithm, the final encrypted cipher text will be,
  uqXRqTULcMqYKwOiyaZnxwfDXMIC89lsmnyTt3orjH+WlfsQp0jlEMF U3CRGZbcv5VkeK0f9E5iANZ+od1hWbxhu9nkX/VagEeSdEgDXQt0k hvPv3U12+nWNdr3TPE38V+hd/j/Jv9bUn8aKf92aID+5bcxWCpbK7Q BNU13LtW9bQN0OhWz4UVTv+HrzVzgbNy0F4g7evOBdxfI4VMqMN W7Wo8uAUTQ4OuT0puPxfms=
- This message is transmitted to the receiver, along with the information about the seeding point to generate the key at the receiver.
- The original plain text is obtained at the receiver side by reversing the encryption process.

V. CONCLUSION:

- The 3-D Pythocrypt algorithm is a new and young technique and is infeasible for many cryptographic attacks.
- Only some part of the numerical values can be obtained and also it is infeasible to find the original plain text by observing the pattern of ciphertext.
- Implementing the 3-D Pythocrypt algorithm along with the AES Encryption algorithm provides better security than many existing cryptographic systems.
- Generating keys using Chitra kavya is a new and powerful technique and it can be used in any scenario where dynamic key generation is required.

VI. ACKNOWLEDGMENT:

We are immensely grateful to Ancient Science and Technology - Center for Incubation, Innovation, Research and Consultancy (CIIRC) and Jyothy Institute of Technology for their encouragement and pearls of wisdom shared during this research.

REFERENCES:

1. H. S. Jois, N. Bhaskar, and M. N. Shesha Prakash, "A 3-d advancement of PythoCrypt for any file type," Journal of Open Innovation: Technology, Market, and Complexity, vol. 1, no. 2, Dec. 2015, DOI: 10.1186/s40852-015-0022-8.
2. D. Yehya and M. Joudi, "AES Encryption: Study & Evaluation Operating Systems View project Logic Design View project." [Online]. Available: https://www.researchgate.net/publication/346446212
3. A. Muhammad Abdullah and A. Muhamad Abdullah, "Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data Application of Petri Nets in Computer Networks View project Call for papers View project Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data,"2017.[Online].Available: https://www.researchgate.net/publication/317615794
4. B. Karthikeyan, K. Alhaf Malik, D. Bujji Babbu, K. Nithya, S. Jafar Ali Ibrahim, N. S. Kalyan Chakravarthy (2021) "Survey of Cooperative Routing Algorithms in Wireless Sensor Networks", Annals of the Romanian Society forCellBiology,pp.5316–5320.Available: https://www.annalsofrscb.ro/index.php/journal/article/view/702
5. Jafar Ali Ibrahim. S, Mohamed Affir. A "Effective Scheduling of Jobs Using Reallocation of Resources Along With Best Fit Strategy and Priority", International Journal of Science Engineering and Advanced Technology(IJSEAT)ISSNNo23216905Vol.2,Issue.2,Feb2014, http://www.ijseat.com/index.php/ijseat/article/view/62

Low – Power Io T-based Weather Monitoring System

Prachi Bhanaria[1]
[1]Department of Electronic Science,
University of Delhi,
South Campus,India
prachibhanariya30@gmail.com,

JitendraSen[2]
[2]ShaktiPumpPvt. Ltd.,
Indore, India
jitendrasen738@gmail.com

RajeevKarothia[3]
[3]Scien techtechnologies
Pvt. Ltd indore, India

Abstract: This paper proposes to monitor weather of specific places on Websites and Android devices using internet of things. The communication between the devices is wireless. The protocol between the units in the design is enhanced to be suitable for most of the appliances. Weather variables such as wind speed and direction, temperature, humidity, solar radiation, soil moisture and rainfall may all be important factors in determining the course of a wide range of events. For example, agriculture has always been heavily dependent on the weather and weather forecasts, both for its control on the quality and quantity of a harvest and its effect on the farmer's ability to work the land or to graze his stock. Water resources generally depend critically not just upon rainfall, but also other weather phenomena that together drive plant growth, photosynthesis, and evaporation. Weather monitoring is also important not just in defining the present climate, but also for detecting changes in climate and providing the data to input into models which enable us to predict future changes in our environment.

Index terms: Internet of Things, Weather Monitoring System, Sensors, Things peaks ever.

1. INTRODUCTION:

Weather forecasting is the application of science and technology to predict the state of the atmosphere for agiven location. Human beings have

attempted to predict the weather informally for millennium and formallysince the nineteenth century. Weather fore casts are made by collecting quantitative data about the current state of the atmosphere on a given place and using scientific understanding of atmospheric processes to project how the atmosphere will evolve on that place. Personal weather stations have the ability to display an as to nishingly large array of weather conditions, whether it is temperature, humidity, wind direction and speed, rain fall, or even solar and ultraviolet radiation.

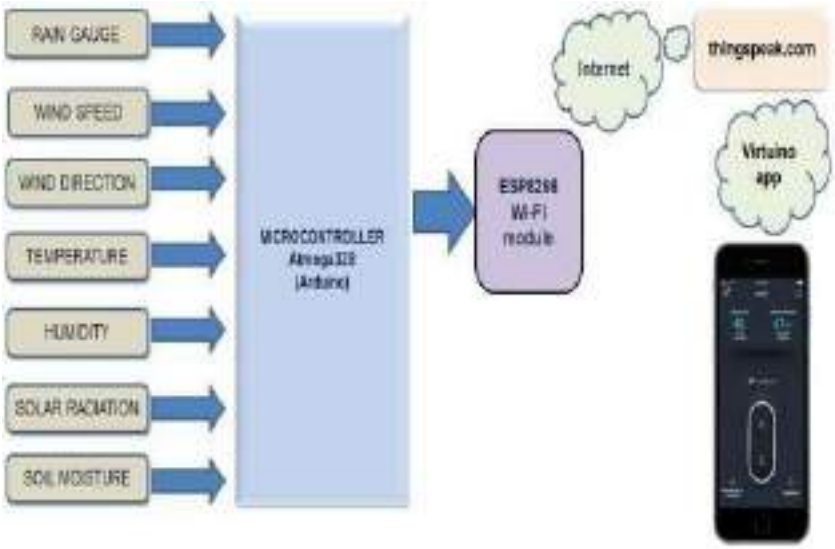
And all of this information can be gathered with out even leaving the house. The weather station provides you with the possibility to share the measured values either by easily integrating it into your own already existing webpage, or you can take advantage of the possibility to use many tools foranalyzing the actual and historical data.

IOT- Based weather monitoring systems are classified based on the technology used. 1) WSN, 2) Satellite, 3) Arduino, 4) GSM, 5) Radar, 6) Zigbee, 7) Camera- Based System. Wireless Sensor Networks (WSNs) [1] include various sensors distributed spatially with the capacity of communication, processing and computing. Here, in real-time manner, data is processed and managed. One proposed framework [2] conquers the above restriction by organization of WSN base for different climate advance utilizing virtual sensor and over la idea.

Fig 1. Weather station

Satellite information is progressively being utilized as a part of conjunction with routine meteorologicalperceptions in the concise

investigation and traditional climate gauge to concentrate data [4]. CanSat [5] is ascale reproduction of the outline, creation and dispatch of a genuine satellite. It is described by minimal effort of usage. Climate observing is the utilization of science and innovation to foresee the condition of the climate for a given area. Through a specific framework [7] [8], it can naturally gather data about stickiness and temperature. The points of interest are put away in a database and as per present and past information authors can deliver the outcomes in a graphical way in the framework. In GSM- based systems[9][10], a gadget forongoing climate observing is displayed to screen the constant temperature, environmental weight, relative dampness, and air's dew point temperature through such system which is utilizing simple and advanced parts. In Radar-based systems like [11] [12], the creators introduced a technique that coordinates both of the information sources to give strategic and arranged climate radar. To create sensor networking and weather station monitoring system without human mediation, utilizing Wireless ZigBee Technology [13][14]. Zigbee is the most recent remote climate checking method. The previous checking frame works of Weather Monitoring System are manual at that time. With a unique sort of camera and computerized multi-image photo grammetric frame work, it's currently conceivable to takeout Digital Elevation Models (DEM) with capturing an image by the camera [15].Using such a strategy; the plane may not be limited to flight way straightly And it region.



2. SYSTEM ARCHITECTURE:

Our wireless weather monitoring system is an automated version of measuring weather parameters and sending the information to a distant database thing speak wirelessly via ESP8266. Our system hasgot almost all things automated so that we get an advantage of this concept, i.e. the real time direct measurement of the parameters through ESP8266. Maintaining back up of sent data is easy and can be done with in a few seconds. This model uses Sensors, ESP8266 module, and a ATMEGA-328 Micro controller.

2.1 At mega 328 based Arduino Board:

AVR RISC-based microcontroller Arduino is both an open-source software library and an open –source break out board for the popular AVR micro-controllers. The Arduino IDE(IntegratedDevelopmentEnvironment) is the program used to write code, and comes in the form ofa downloadable file on the Arduino website. The Ard \uino board is the physical board that stores and performs the code uploaded to it. Both the soft ware package and the board are refer red toas" Arduino.

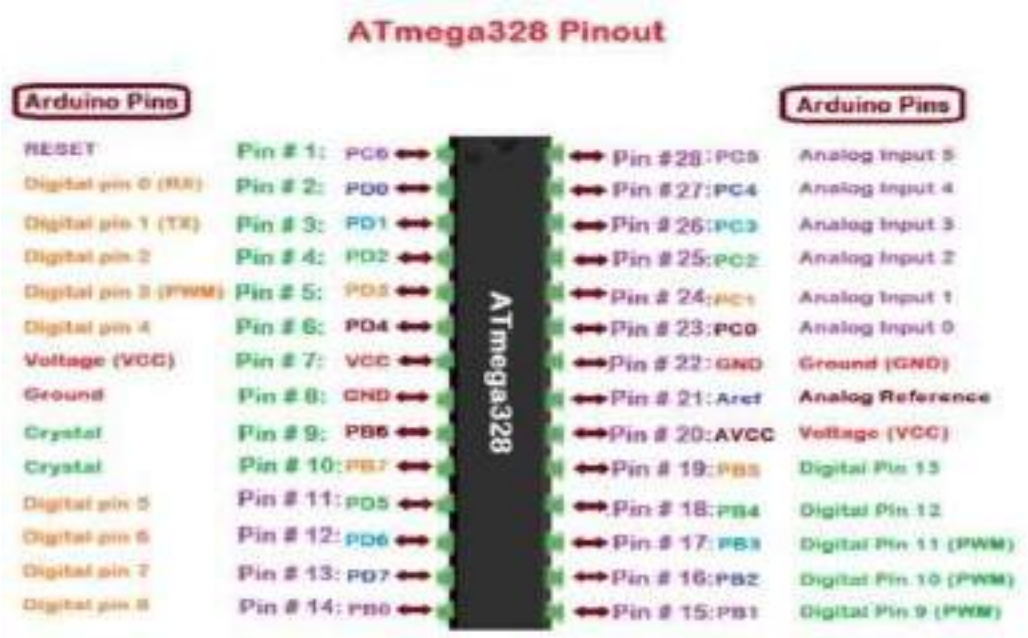


Fig3.Atmega-ArduinoPinmapping

2.2 ESP8266 Wi-Fi module:

The ESP8266 is a Wi-Fi SOC (system on a chip) produced by Espress if Systems. The ESP8266 can be used as an external Wi-Fi module, using the standard AT Command set Firmware by connecting ittoany microcontroller using the serial UART, or directly serve as a Wi - Fi enable dmicro-controller, by programming a new firmware using the provided SDK. This board has been around for almost a year now ,and has been used mostly in IOT contexts, where we want to add connectivity, for example to an Arduino project.

Fig4. ESP8266Wi-Fi Module

2.3. Sensors :

2.3.1. Wind Speed (Anemometer):

An anemometer is a device used for measuring the speed of wind, and is also a common weather station instrument. The wind moves the cups on the anemometer, which in turn rotates an enclosed magnet. The magnet closes a reed switch on each rotation, which is reflected on the output. You can measure this on the two innerconductorsoftheRJ-11 connector (pins2and3), using a digital counter micro controller.

Fig5. Wind Speed Sensor

2.3.2. Wind Direction:

The wind vane indicates the direction that the wind is blowing. This is actually the most complex gauge of the three. Internally on the vane are eight switches, each with their own unique resistor. As the wind be rotates, a magnet closes the reed switches, and may close two at a time due to their proximity to each other. With the use of an external resistor, a voltage divider can be created.

Fig5. Wind Direction Sensor

2.3.3. Rain Guage:

The rain gauge measures rain fall. The sensor is a self-emptying tipping bucket collector. This means that for each 0.011" (0.2794 mm) of rain that falls on the sensor, the bucket will tilt, dumping the water out and closing a momentary contact. The closure of the momentary switch can be measured using interrupt pins or a digital counter. The center conductors of the RJ-11 connect or are connected to the gauges switch.

Fig6. Rain Guage Sensor

2.3.4. Solar Radiation sensor:

The sensor is designed for continuous measurement of solar radiation or light measurements in open field forlong durations. The low power consumption makes it ideal for remote applications. The sensor provides alinear0 to 5 volt output.

Fig7. Solar Radiation Sensor

2.3.5. Soil moisture sensor:

Soil moisture sensors measure the volumetric water content in the soil. Since the direct gravimetric measurement of free soil moisture requires removing, drying, and weighting of a sample, soil moisture sensors measure the volumetric water content indirectly by using some other property of the soil, such as electrical resistance, dielectric constant, orinter action with neutrons, as a proxy for the moisture content.

Fig8. Soil Moisture Sensor

2.3.6. Temperature & Humidity sensor:

A temperature sensor is a device that provides temperature measurement through an electrical signal. A humidity sensor senses, measures and regularly reports there relative humidity in the air. It measures both moisture and air temperature.

Fig9. Temperature and Humidity Sensor

2.3. Printed circuit board:

A Printed Circuit Boards (PCB) is a rugged, copper and non-conductive substrate-based structure used to connect electrical components. The PCB is the backbone of electrical devices, allowing you to connect passive (resistor, inductor, capacitors, etc.), active (operational amplifiers etc.) and embedded devices together, into specific form factors to fit the design need. Connections between the components are made through copper connections (routes) which become passageways for electrical signals.

3. SIMULATION SAND EXPERIMENTAL RESULTS:

The simulation is done on Arduino IDE. The output voltage of sensors is obtained in milli volts and is converted to digital value. The Hardware system provides the weather information through the Thing speaks ever. The ESP8266 module works in station mode so we provide a connectivity through any router, then this module sends all collective sensors values to thing speak.

All sensors are connected to port of the ATMEGA 328 micro co ntroller device. The over all power consumption by the system is very less

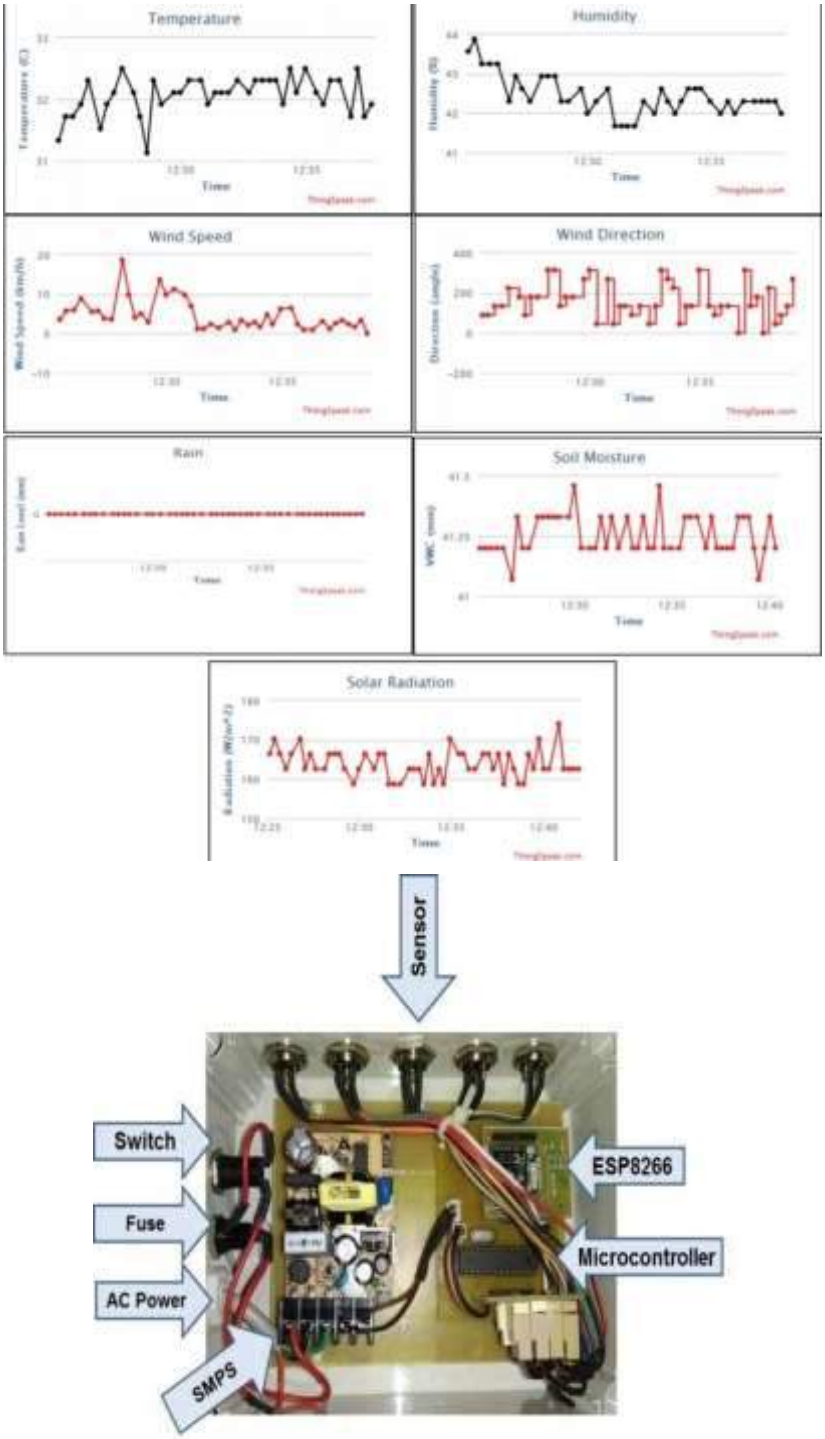


Fig11. Final result on things peak System

Table1. Power Consumption of components

Components	Power(approx.)
Microcontroller	80mW
ESP8266	450mW
RainSensor	5mW
WindSpeed Sensor	5mW
WindDirectionSensor	5mW
Temperature andHumiditySensor	10mW
SoilMoistureSensor	30mW
SolarRadiationSensor	150mW
Total Power	730mW(at 5vand140 mA)

All collected values can be monitored directly which is simultaneously displayed on the Virtuino app paneland a message is sent to the mobile by using Virtuino when some particular condition is occur at the same instance.

4. CONCLUSION:

The paper deals with designing a weather monitoring system using some sensors, ESP8266 module andATmega328 microcontroller unit to monitor weather conditions of the desired location and transmit it to a web server and android application. The designed product module is at the prelim stage and designed for monitoring some parameters like Temperature, humidity, rain level, wind speed, etc but can be enhanced for monitoring other different types of environmental and climatic behavior of a location, which also can consume less power. In this paper we use the ESP8266 Wi-Fi module for sending data to the server so it requires connectivity with any router when connectivity is lost, the information will not sendby the station for avoiding this we can also use in built connectivity by using GSM module.

5. ACKNOWLEDGMENT:

This work was funded by the Scientech Technologies Pvt. Ltd. Indore and supervised by Mr. Rajeeev Karothia, Department of Research & Development.

REFERENCES:

1. Amber Katyal, Ravi Yadav, Manoj Pandey, Wireless Arduino Based Weather Station, Journal of Computer Science,Vol.5, Issue4, April 2016
2. Andriy Holovatyy,Development of IoT Weather Monitoring System Based on Arduino and ESP8266 Wi-Fi Module, CAD in Machinery Design: Implementation and Educational Issues (CADMD), 2020
3. B. Karthikeyan, K. Alhaf Malik, D. Bujji Babbu, K. Nithya, S. Jafar Ali Ibrahim, N. S. Kalyan Chakravarthy, "Survey of Cooperative Routing Algorithms in Wireless Sensor Networks", Annals of the Romanian Society for Cell Biology, pp. 5316–5320, 2021 https://www.annalsofrscb.ro/index.php/journal/article/view/702
4. D. S. Arjun, A. Bala, V. Dwarakanath, K. S. Sampada, B. B. Prahlada Rao and H. Pasupuleti, Integrating cloud-WSNto analyze weather data and notify SaaS user alerts during weather disasters, IEEE International Advance ComputingConference(IACC),pp. 899-904, 2015.
5. Gopal G, Harith B, Ritwik Raj Savyasachi ChetanUmadi, Weather Monitoring Using Parachute Satellite Can Sat, International Journal of Engineering Science and Computing, Volume 6 Issue, May 2016
6. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IOT Sensor Networks." Advances in Information Communication Technology and Computing, pp. 461-471. Springer, Singapore, 2022 https://link.springer.com/chapter/10.1007/978-981-19-0619-0_41
7. Karthik Krishnamurthi, Suraj Thapa, Arduino Based Weather Monitoring System,JournalofComputerScience,Volume3,Issue2, March-April 2015
8. KC Gouda, Preetham VR, MN Shanmukha Swamy, Microcontroller Based Real Time Weather Monitoring Device With GSM, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 7, July 2014
9. Keshav kumar Singh, Design of Wireless Weather Monitoring System using GSM, Thes is submitted to Department of Electronics & Communication Engineering National Institute of Technology Rourkela, 2013

10. K. Sai Nikhilesh, Y. Harish Raaghavendra, P.J. Madhu Soothanan, Resmi R, Low-cost IoT based weather monitoring system for smart community, Proceeding soft he Fourth International Conference on Inventive Systems and Control (ICISC), 2020
11. Nisha Gahlot, Varsha Gundkal, Sonail Kothimbire, Archana Thite, Zigbee based weather monitoring system, International Journal Of Engineering And Science (IJES), Volume4 Issue4. ISSN(p): 2319–1805, 2015
12. Rajdeep Kumar, Sazid Khan, Manik Chand Yadav; S K Dubey, Weather monitoring system, International Journal of Advanced Technology in Engineering and Science Volume 02, Issue No. 05, May 2014
13. S.B.Kamble, P.Ramana P.Rao, AnuragS. Pingalkar, GaneshS. Chayal, IoT Based Weather Monitoring System, IJARIIE-ISSN(O) -2395 -4396, Vol-3Issue-22017
14. Srinivasa K.G. M.S.Ramaiah. Siddiqui.N. Kumar. A,ParaSense - A Sensor Integrated Cloud Based Internet of Things Proto type for Real Time Monitoring Applications, inregion 10IEEE Symposium (TENSYMP),2015.

Automated Health Monitoring System for Premature Fetus

Dr. S. Jafar Ali Ibrahim[1], D. Hemalatha[2], V. Pavan Kalyan[3], K. Dinesh chandra[4],M. Harsha Vardhan[5]

[1]Associate Professor, Dept. of Information Technology, QIS College of Engineering and Technology, AndhraPradesh, India

[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, AndhraPradesh, India

Abstract - Our research is about the premature infant Monitoring system based on wireless Technology. A prototype is developed which gives a reliable and efficient baby monitoring system that can play a significant role in providing higher kid care. This system monitor vital parameters such as body Temperature, Pulse rate, Cry Movement,Blood oxygenlevel(spo2),heart rate and the physical activity of an infant and using GSM Network.GSM is an open and digital cellular technology used for transmitting mobile voice and data services operates at the 850MHZ and 1900MHZ frequency bands. This information is transferred to their parents. Measurements of these vital parameters can be done and underneath risk, Scenario sent to the parents with SMS alert to intimate the proper control action. So that we can reduce the death of the premature fetus.

Keywords- NICU , Preterm, Incubators, Infants

.

1. INTRODUCTION:

A device is developed which gives a reliable and efficient baby monitoring system that can play a vital role in providing better infant care. This device monitor vital parameters such as body temperature, heartbeat, cry, movement of an infant and sends SMS alert using wireless technology. Our proposed system aims at monitoring the vital signs of the premature infant such as heartbeat, movement of body, body temperature and cry using wireless technology and sensors which are comfortable for the baby to wear. It is also accurate and precise than other sensors. We also focus on increase the scope

of transmitting the information without over the internet in orderto provide remote access**.** This system overcomes the drawback of the existing systems which are clumsy, less user friendly and expensive may cause discomfort to the infant. Wireless and wearable sensors provide more convenient and long term monitoring. Sudden Infant Death Syndrome (SIDS) is the unexplained death of an infant below the age of one year. It usually happens without any warning signs during sleep, which is why it is difficult to identify and predict. Therefore our proposed monitoring system would be an effective way to predict the onset of SIDS .Neonatology is a subspecialty of pediatricks that started to develop in the 1940s. After the World War–II the specific needs of sick newborn infants were recognized and new prematurenurseries were built. The term "Neonatology" was first used by Alexander Schaffer in 1960 in the introduction of the first edition of his book.

Miniaturization of samples for blood tests, needed for clinical management including electrolytes, bilirubin and blood gases was one of the major advances in the development of Neonatology. In the following decades important progress was achieved in thermoregulation, nutrition, growth, respiratory support, cardiopulmonary support and infection control.

2. LITERACTURE REVIEW:

There are many designs of incubator for infants in the literatures. In recent work, Dive and Kulkarni designed an incubator that can monitor and detect the light inside the incubator, and also audio or voice of the infant . The proposed incubator system can notify doctor and nurse about the infant's condition, as when the infant cries, the alarm will be triggered and the alarm will stop or deactivated only if someone turned it off. The advantage of the work is it helps doctors and nurses to monitor the infant's condition continuously. For future improvement, they recommended adding parameters such as monitoring of Temperature and moisture, developed a newborn incubator that can check the conditions of the incubator environment by utilizing a humidity control system. They concluded that the control of humidity could contribute to the thermos-neutral of the environment, thus improving the premature newborns' quality of life. There are also several infant

incubator designs that implement temperature control system. However, following care giving, infant and incubator temperature differed significantly over time by incubator control mode (air mode control or skin temperature mode control) . Therefore, it is necessary to consider the temperature effects of care giving when developing incubators. There are several others unresolved issues in developing infant incubators such as exposure to high noise levels in NICU, incubator's surrounding light environment and electromagnetic fields (EMFs) impact on infant health to name a few.

Based on the literature review the gaps are identified with the Existing system regarding the Sensors so that we developed proposed system.

.

3. PROPOSED SYSTEM:

Above, monitoring of moisture level and measurement of infant's temperature are important parameters to be considered in designing an infant incubator. Furthermore, an alarm system that can detect both parameters are nearing safety level threshold, and then, alert caregivers about this situation is also essential to prevent harmful situation. Therefore, the proposed infant monitoring system will consist of an alarm system and sensors that can measure incubator's monitor temperature.

4. SYSTEM ARCHITECTURE:

The architecture of the system consists of both hardware and software. Block diagram hardware components were assembled according to the block diagram. The code is written in embedded C and is burnt into the ARDUNIO UNO R3

Fig.1. Block Diagram of Proposed System

5. COMPONENTS USED:

A.TEMPERATURE SENSOR:

Human body needs special type of sensors for reliable readings which led to the choice of using the LM35 temperature sensors in our prototype . It operates at 3 to 5 V and can measure temperature in the range of - 40 C to +125 C which issufficient for the targeted body temperature range.

Measure incubator's monitor temperature

B. PIR MOTION SENSOR :

A PIR (Passive Infra Red) sensor is a motion detector which detects the heat (infrared) emitted naturally by humans .When a person in the field of vision of the sensor moves, the sensor detects a sudden change in infrared energy and the sensor is triggered (activated).They are commonly used in security lighting and alarm systems in an indoor environment. The PIR sensors have a range of approximately 6 meters, depending on conditions. The sensor adjusts to slowly changing conditions that occur normally within the environment, but shows a high-output response when a sudden change takes place.

Fig.2. PIR sensor

C. PULSE RATE SENSOR :

A Pulse rate sensor is a monitoring device that allows one to measure the infant's heart rate in real time . It provides a simple way to study the heart function. This sensor monitors the flow of blood through the finger and is designed to give digital output of the heartbeat when a finger is placed on it. When the sensor is working, the beat LED flashes in unison with each heartbeat.

D. SOUND DETECTION SENSOR:

The sound sensor module provides an easy way to detect sound and is generally used for detecting sound intensity. This module shown as in fig can be used for security, switch, and monitoring applications. Its accuracy can be easily adjusted for the convenience of usage. It uses a microphone which supplies the input to an amplifier, peak detector and buffer. When the sensor detects a sound, it processes an output signal voltage which is sent to a microcontroller then performs necessary processing

IV. ADRUINO UNO R3 BOARD

The Arduino UNO is a widely used open-source microcontroller board based on the Atmega 328P microcontroller and developed by Arduino.cc. [The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards

(shields) and other circuits . The board features 14 Digital pins and 6 Analog pins. It is programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable. It can be powered by a USB cable or by an external 9 volt battery, though it accepts voltages between 7 and 20 volts. It is also similar Leonardo.

Fig.6. Arduino UNO R3

4. EXISTING SYSTEM:

- **Temperature Sensor** – Human body needs special types of sensors reliable readings which lead to the choice of using the LM35 temperature sensors in our property.
- **Sound Cancelling** – Design and Development of a smart baby monitoring system based on raspberry pi and pi camera.
- Hence there are a lot of gaps to rectify about the sensors of Temperature, heart beat and etc.,we developed a latest sensors which gives alaram signals to the parents so that we can take care of babies.

6. RESULT:

The system was tested carefully on an infant, the results found to be same as the one's measured by standard instrumen.While testing this system on an infant parent's concern was considered. During the execution of the system snapshots of the display were taken. The system being a complete hardware design and the data available on cell phone.

TABLE.I

Serial No	Actual Temp (0C)	Practical Temp(0C)
1	32	36.1
2	31	35.5
3	33	37
4	35.6	36.7

TABLE.II

Serial No	Actual pulse rate	Practical pulse rate
1	72	78
2	66	72
3	70	76
4	54	60

7. CONCLUSION :

There are an increase number of premature babies that, once they achieve some degree of maturity, can be moved at home when their parents can take care of them in a loving environment which most of the times fasten the babies' recovery. For this end, a home based solution is provided, which supports parents in the baby care while keeping babies under the control of clinical staff. The architecture contains two case-based reasoning modules that assess the baby state. Then, the recommendations and alerts obtained are shown in parents' mobile as well as in doctor's device. The proposal of a general and modular architecture enables that this solution can be applied to other use cases within the Moshca platform. Even the reasoning module can be replaced by other module using different artificial intelligent methods according to the current necessities. Moreover, the use of recognized standards worldwide ensures the interoperability of the proposed architecture. Our future work involves further experimentationincluding the testing in a real environment and the comparison of results obtained by other techniques. In addition, the influence of context awareness on obtained recommendations should be studied.

8. REFERENCES:

1. J. Bunker, M. Kejariwal and G. Monlux(Oct 28-31, 1993), "SIDS home monitor with telecommunications capabilities", In Proc. of IEEE EMBS International Conference, pp. 1060–1061.
2. Intelligent Baby Monitoring System 1Savita P.Patil, 2Manisha R. Mhetre, ITSI Transactions on Electrical and Electronics
3. Engineering (ITSI-TEEE) ISSN (PRINT) : 2320 – 8945, Volume -2, Issue -1,2014.
4. J.E.Garcia, R.A.Torres, "Telehealth mobile system ", IEEE Conference publication on Pan American Health Care Exchanges, May 4,2013.
5. Karthick, R and Sundararajan, M: "A Reconfigurable Method for Time Correlated Mimo Channels with a Decision Feedback Receiver," International Journal of Applied Engineering Research 12 (2017) 5234.
6. Karthick, R and Sundararajan, M: "PSO based out-of-order (ooo) execution scheme for HT-MPSOC"Journal of Advanced Research in Dynamical and Control Systems 9 (2017) 1969.
7. Karthick, R and Sundararajan, M: "Design and Implementation of Low Power Testing Using Advanced Razor Based Processor," International Journal of Applied Engineering Research 12 (2017) 6384.
8. Karthick, R and Sundararajan, M: "A novel 3-D-IC test architecture-a review," International Journal of Engineering and Technology (UAE)7 (2018) 582.
9. R.Karthick, P Selvaprasanth, A ManojPrabaharan, "Integrated System For Regional Navigator And Seasons. Management," Journal of Global Research in Computer Science 9(4),2018(11-15).
10. R. Karthick,NSathiyanathan, and M. Eden, "Medical Image Compression Using View Compensated Wavelet Transform" Journal of Global Research in Computer Science 9(9),2018(1-4).
11. Karthick, R and Prabaharan, A.Manoj and Selvaprasanth, P. and Sathiyanathan, N. and Nagaraj, A., High Resolution Image Scaling Using Fuzzy Based FPGA Implementation (March 15, 2019). Asian Journal of Applied Science and Technology (AJAST), Volume 3, Issue 1, Pages 215-221, Jan-March 2019 . Available at SSRN: https:// €ssrn.com/abstract=3353627

12. Karthick, R and Sundhararajan, M., Hardware Evaluation of Second Round SHA-3 Candidates Using FPGA (April 2, 2014). International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014), Vol. 2, Issue 2, Ver. 3 (April - June 2014). Available at SSRN: https://ssrn.com/abstract=3345417.
13. Karthick, R and andPrabaharan, A.Manoj and Selvaprasanth, P.,Internet of Things based High Security Border Surveillance Strategy (May 24, 2019). Asian Journal of Applied Science and Technology (AJAST), Volume 3, Issue 2, Pages 4-100, Apr-June 2019. Available atSSRN: https://ssrn.com/abstract= 3394082.
14. Karthick, R and Sundararajan, M: "SPIDER based out-of-order (ooo) execution scheme for HT-MPSOC" International 11.M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong,pp352-353 http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf
15. C. Narmatha , Dr. M. Thangamani , S. Jafar Ali Ibrahim, " Research Scenario of Medical Data Mining Using Fuzzy and Graph theory", International Journal of Advanced Trends in Computer Science and https://doi.org/10.30534/ijatcse/2020 https://www.warse.org/IJATCSE/static/pdf/file/ijatcse52912020.pdf
16. Faxin Yu, Zheming Lu, Ha-o Luo, Pinghui wang, "Three dimensional Model analysis and processing", pp.374-378, 2011.
17. Celik Mehmet Utku, Gaurav Sharma, Ahmet Murat Tekalp, Eli Saber, "Lossless Generalized-LSB Data Embedding", IEEE Transactions on Image Processing, vol. 14, no. 2, pp.253-266, Feb. 2005
18. Wien Hong, "Adaptive reversible data hiding method based on error energy control and histogram shifting", Elsevier, Optics Communica 2012.
19. Xiaolong Li, Weiming Zhang, XinluGui, and Bin Yang, "A Novel Reversible Data Hiding Scheme Based on TwoDimensional Difference-

Histogram Modification", IEEE Transactions on Information forensics and Security, vol. 8, no.7, pp.1091-1100, July 2013.

20. Keren Wang, Hong Zhao and Hongxia Wang, "Video Steganalysis Against Motion Vector-Based Steganography by Adding or Subtracting One Motion Vector Value", IEEE Transactions on Information Forensics and Security, vol. 9, no. 5, pp. 741-751, May 2014.
21. Gallaiszx A, Hedli T-H, Loscrix V, Mitton N (2019) Denial-of-sleep attacks against IoT networks. In: 6th international conference on control, decision and information technologies
22. Thangamani M., and S. Jafar Ali Ibrahim. Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso -Wc) and Gene Ontology for Microarray Gene Expression. In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, ACM pp. 48-55. 2018. https://dl.acm.org/doi/abs/10.1145/3299852.3299866
23. Ibrahim, S.jafar Ali and M.Thangamani. Prediction of novel drugs and diseases for hepatocellular carcinoma base of multi-source simulated annealing based random walk, Journal of medical systems,42, no.10 (2018):188.https://doi.org/10.1007/s10916-018-1038-y

E - Learning Management System

Jeevan N
Department of Information Science and Engineering
Sai Vidya Institute of Technology Bangalore, India
jeevann.18 is @ saividya. ac.in

Vadiraja H S
Department of Information Science and Engineering
Sai Vidya Institute of Technology Bangalore, India
vadiraj ahs.18 is @ saividya. ac.in

Nithish Baradwaj V
Department of Information Science and Engineering
Sai Vidya Institute of Technology Bangalore, India
nithish baradwaj baradwajv.18 is @ saividya.ac.in

Prof. Santhosh Y N
Department of Information Science and Engineering
Sai Vidya Institute of Technology Bangalore, India
santhosh. Yn @ saividya. ac.in

V Adithya
Department of Information Science and Engineering
Sai Vidya Institute of Technology Bangalore, India
adithyav.18 is @ saividya. ac.in

Abstract: The goal of the E-learning Management System is to automate the existing manual system using computerized equipment and full-featured computer software that meets their needs, so that their valuable data/information can be stored for a longer length of time and manipulated easily. The required software and hardware are both easily available and straightforward to use. An E-learning Management System, as previously said,

can lead to error-free learning. System of management that is safe, dependable, and quick It might help the user focus on their other tasks rather than maintaining track of their records. As a result, it displays how user-friendly it is. Without redundant ententes, the business can keep digital records. That is to say, in order to access the Information, one does not need to be distracted by irrelevant information. The goal is to replace their current manual system with automated equipment and full-featured computer software to meet their needs, so that their valuable data/information can be saved for a longer period of time with easy access and manipulation Essentially, the project demonstrates how to manage for improved performance and client service.

Keywords: component, formatting, style, styling, insert (keywords)

INTRODUCTION:

An E-learning management system is a software application that is used to distribute, administer, and record any educational content such as courses, training models, and other educational content for a student's better understanding and learning Approximately 3/4 of the overall learning system market is made up of E-LMS.. E-LMS were created to fill in the gaps in terms of teaching a student with the use of analytical data and reporting. E-LMS delivers a wide range of audiences with online data content. Higher education students use the E-LMS as a virtual classroom. An LMS often allows an instructor or teacher to generate and deliver curriculum, track student involvement, and evaluate student performance Learning management systems are used by businesses of all sizes, national government agencies, local governments, traditional educational institutions, and online/eLearning-based institutions. The systems have the potential to improve traditional instructional techniques while also saving time and money for companies. With a good system, instructors and administrators will be able to manage user registration, material, calendars, user access, communication, certifications, and notifications more properly. To eliminate data entry errors, the programme is kept as simple as feasible. When entering invalid data, it also displays an error notice. The user does not require any formal

knowledge to use this system. As a result, it demonstrates that it is user-friendly. As previously said, an E-Learning Management System can lead to an error-free, secure, dependable, and rapid management system. It might help the user focus on their other tasks rather than maintaining track of their records. As a result, the company will be able to make better use of its resources. In this E-Learning system you will be able to add student in a class, upload file, add Course, Department, Subject. It also has a form validator and a Responsive Design that works on both phones and tablets. This project has a lot of features that are easy to maintain. We also supply a PHP-based e-learning project report. This project includes a large number of advanced modules, resulting in a very powerful back end system. In this project, we created a large number of modules.

Problem Statement:

- To address the issues with the current system, such as one-on-one engagement, proper note distribution, and a lack of proper teamwork.
- The current system is deficient in key elements.
- In terms of functionalities, the current system is lacking.
- To incorporate features such as quizzes and blanks.
- To aggregate third party notes and videos

A. Objective:

The major goal of the E-learning Management System Project is to keep track of the details of assignments, students, teachers, quizzes, and questions. It keeps track of all information pertaining to Assignment, CLASS, QUESTION, and Assignment. Only the administrator gets access to the project because it is totally built at the administrative level. The project's purpose is to develop software that will reduce the amount of human labor required to manage Assignments, Students, Classes, and Teachers. It keeps track of all of the information regarding the TEACHER, QUIZ, and QUESTION.

B. Scope of the Project:

It could be beneficial for gaining complete information about perfect management. In a short period of time, the collection will be obvious, simple,

and sensible. It will assist a person in comprehending the management of the previous year in a clear and vivid manner. It also helps to finish all present E-learning Management System initiatives. The cost of collecting the management will be reduced, and the XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE collection process will be more efficient. Because the goal of our project is to automate business operations, we've sought to computerize a number of E-learning Management System processes. A user needs fill out several forms on a computer system, and a huge number of copies of the forms can be generated swiftly.

- In a computer system, it is not essential to create the manifest; instead, we can print it directly, saving time.
- Assisting staff in documenting their job efforts in various departments.
- To maximise resource utilisation by increasing productivity through automation.
- The system generates a wide range of data that can be applied to a number of tasks.

RELATED WORK:

(a) Zoom:

Zoom is a cloud-based video conferencing service that allows you to virtually meet with others through video, audio, or both while conducting live discussions - and it also allows you to record such sessions for later viewing.

Features:

- Screen-sharing to present documents, spreadsheets, presentations, or (if using a browser) other browser tabs
- Two-way and multi-way audio and video calls
- An accompanying chat Disadvantages: ❑
- Attendees can't search for their interesting event or meeting and join
- Only if the link/Meeting ID is present with the user only then he can attain the event or the meeting.

(b) Edmodo:

Edmodo is an educational technology company that provides K-12 schools and instructors with a collaboration and coaching platform. Teachers can use the Edmodo network to exchange information, assign quizzes, and manage contact with students, colleagues, and parents.

Features:

- Facilitate learning goals.
- Create polls for students.
- Online classroom discussions.
- Assess student progress.Disadvantages:
- No proper face to face interaction.
- Meeting conduction is not possible.

(c) Microsoft Teams:

Microsoft Teams is a persistent chat-based collaboration tool that includes document sharing, online meetings, and a slew of other business-friendly capabilities.

Features:

- Call encryption between all users
- Screen-sharing to present documents, spreadsheets, presentations, or (if using a browser) other browser tabs
- Two-way and multi-way audio and video calls
- An accompanying chat.
- Ability to raise and lower handDisadvantages:
- Attendees can't search for their interesting event or meeting and join
- Only if the link/Meeting ID is present with the user only then he can attain the event or the meeting

(d) Byju's:

- BYJU'S is an education tutoring app that operates on a freemium model, with free content access limited to 15 days following registration. It first opened its doors in August 2015, with instructional content for kids in grades

4 through 12, and in 2019, an early learning programme for grades 1 through 3 was created. Students are also prepared for national exams such as IIT-JEE, NEET, CAT, and IAS, as well as international tests such as the GRE and GMAT.

Features:

- Lessons on video. BYJU'S provides video classes andlectures on a variety of topics, as well as personalizedinstruction from educators.
- Personalization
- Mapped to the Syllabus
- Unlimited practice
- In-Depth Analysis.
- Questions Library.Disadvantages:
- The only con is that Byjus is really expensive.

(e) Google Classroom:

Google Classroom is a free blended learning software for educational institutions that promises to make creating, sharing, and grading lessons easier.

Features:

- Start a meeting.
- Without using a pen and paper, create and manage class rooms and assignments online.

Disadvantages:

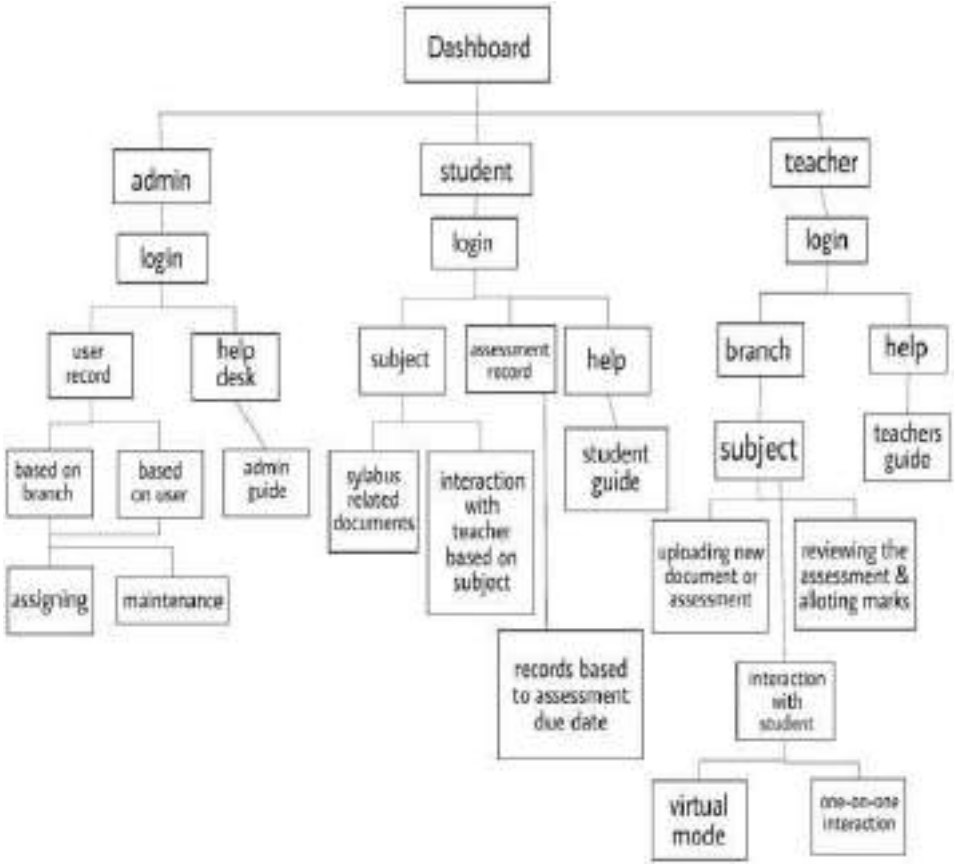
- Limited integration options.
- Not a video solution.

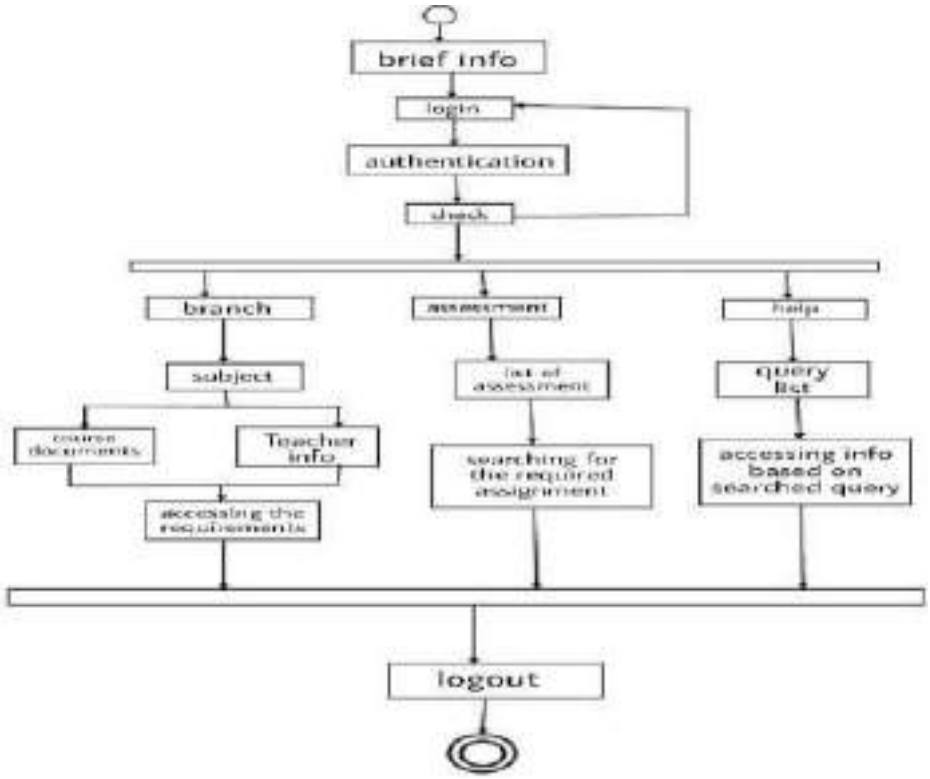
PROPOSED WORK:

The suggested system's goal is to provide a system with better facilities. The proposed solution can overcome all of the existing system's flaws. The method ensures sufficient security while also reducing manual labor.

- Security of data.
- Make sure the data is correct.
- Appropriate control of higher-ranking officials.
- Minimize data entering by hand.
- Minimum time needed for the various processing. User friendliness and interactive.
- Minimum time required.

ARCHITECHTURE:





CONCLUSION:

- This application automates the process of studying and self-evaluating online courses.
- It is used by students to attend live sessions.
- Among other things, it's utilised to deliver and manage movies, papers, courses, and quizzes.
- This project's execution benefits both students and teachers.

It may also be used by any management to improve any of their assignments on a regular basis in any firm. For example, in terms of project success, an X firm manager can allocate time-to-time assignments to their personnel.

REFERENCES:

1. https://www.researchgate.net/publication/283331228_Acation of Learning Management System for Online Learning Modules

2. https://www.researchgate.net/publication/309242990 A _Literature on E Learning Systems in Higher Education
3. https://www.researchgate.net / publication/ 327417783_ Effectiveness of Google Classroom Perception
4. https://en.wikipedia.org/learning_management system
5. https://ieeexplore.ieee.org/document/4281114/
6. https://ieeexplore.ieee.org/document/5993711/
7. https://ieeexplore.ieee.org/document/5496415/
8. https://ieeexplore.ieee.org/document/7569542

Fuzzy Dbg : An integrated debugger and Fuzzer

[1]Sai RitvikTanksalkar, [1]Sandhra Bino, [1]Vishnu Madhav, [1]Vivek Kamisetty, [1]Jayashree Narayanan, 2Vipin Pavitra

Department of Computer Science and Engineering 2Amrita Center for Cyber security Systems and Networks

Amrita Vishwa Vidyapeetam, Amritapuri, India

{sairitvikt ,sandhrabino, pratapagiri vishnumadhav, kamisettyvivek @ am. students. amrita.edu, {jayasreen, vipinp)@ am.amrita.edu

Abstract : Generic debuggers like GDB neither have functionalities like viewing execution traces when a crash or an unintended event occurs or an efficient fuzzing functionality built into the heart of the debugger itself. Fuzzing is the most advanced and fastest way to find bugs in software. This project aims to build a simple and fast debugger with an implementation of execution trace and fuzzing functionality. We have implemented fuzzing as a module of the debugger, the fuzzer generates mutated inputs based on an initial seed and saves the crashing input inside another directory for future references.

Keywords: Fuzzing, Execution Traces, Mutation, Debugger

I. INTRODUCTION:

Analyzing soft wares to resolve crashes or fix vulnerabilities has always been an exigent task. Most of the proprietary soft wares don't even provide source code for static analysis. Application debugging has become an inevitable part of any software development cycle. It is increasingly important in modern day programming practices because of the growing complexity of software and hardware [1]. In cases where black box applications are involved, fuzz testing becomes the need of the hour to find and fix critical security as well as non-security vulnerabilities. Fuzzers produce a large number of inputs in a small amount of time to test the program in a short time. The fuzzer module of the debugger discussed in this paper aims to create such mutated inputs to feed to binaries to detect basic software security vulnerabilities like buffer overflows.

Program execution traces provide the most intimate details of a program's dynamic behavior. They can be used for program optimization, failure diagnosis, collecting software metrics like coverage, test prioritization, etc. Two major obstacles to exploiting the full potential of information they provide are: (i) performance overhead while collecting traces, and (ii) significant size of traces even for short execution scenarios [2]. The tracer engine module of the debugger discussed in this paper aims to reduce the aforementioned performance overhead while collecting and displaying traces by scanning for static symbols from the debugee and dynamic symbol addresses from the system's standard library.

Most modern debuggers neither provide a fuzzer integrated right inside the debugger nor an execution trace to analyze what important functionalities have been triggered prior to triggering a crash. This is where Fuzzy Dbg comes into picture - it provides a fuzzer as well as an execution tracer engine along with generic debugging capabilities in an attempt to achieve a one stop debugging solution for all soft wares.

II. RELATED WORK:

Mainstream debuggers like GDB offer the possibility to suspend the program execution by placing breakpoints. When the program execution arrives at a breakpoint, the execution is suspended and the debugger shows to the developer the current state of the execution. This execution is generally shown as the state of registers, disassembly at the breakpoint, and a stack trace [3]. From a security perspective, particularly for exploit developers and reverse engineers, it is sometimes required to analyze the set of functions which were triggered during the execution which ultimately caused the crash. In such analysis cases, an execution trace is preferred more than the stack trace. Linux kernel specifically provides the ftrace utility to do exactly this [4], but the aim of our paper is to achieve a proper execution trace with less implementational convolution and inside the debugger itself.

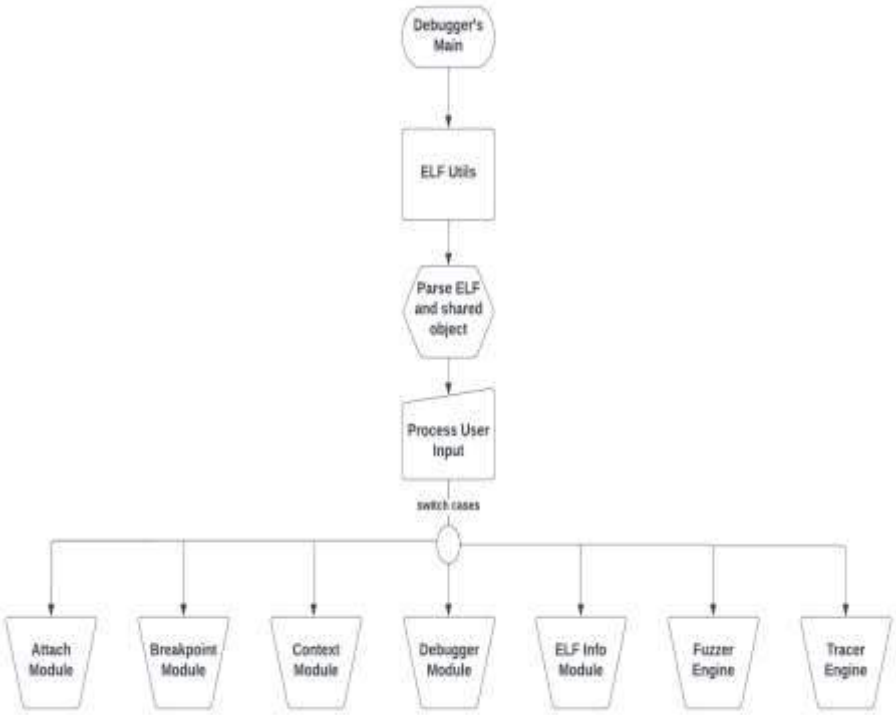


Fig.1 Implementational flow of the system

Fuzzing these days is one of the most extensively used vulnerability detection techniques. A great many fuzzers like AFL, syzkaller, clusterfuzzetc all have their specific uses for fuzzing various types of applications ranging from internet protocols to browsers to kernels. All these fuzzers are standalone tools used for various scenarios. Our fuzzer module comes integrated right into the debugger itself. The module relies on mutation based fuzzing, generating inputs based on a user provided seed. The fuzzer is designed to keep track of all the seeds which trigger a crash for future references.

III. SYSTEM OVERVIEW:

The entire system is divided into different modules. The Fig.1 shows the implementational flow of the entire system broken down into several modules for the sake of brevity. Most of the modules implemented are independent of all other modules and function asynchronously.

01. Attach:

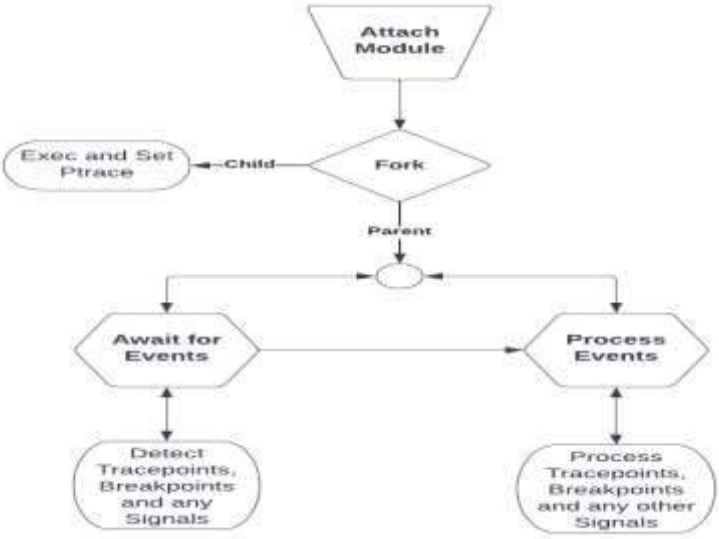
The main aim of this module is to attach the debugger to the entry point of the debugee. The internal implementation consists of forking the debugee into a child process. The ELF as well as the shared library of the system are parsed to extract static and dynamic symbols respectively. Trace points are set at every function and appropriate handlers are set to capture the trace points. It then uses ptrace to communicate and instrument the child process for debugging. Appropriate events like breakpoints, signals are all captured by the event handlers and processed. Every event is assigned a special structure as can be seen from Fig.4 to handle various events and their respective return values.

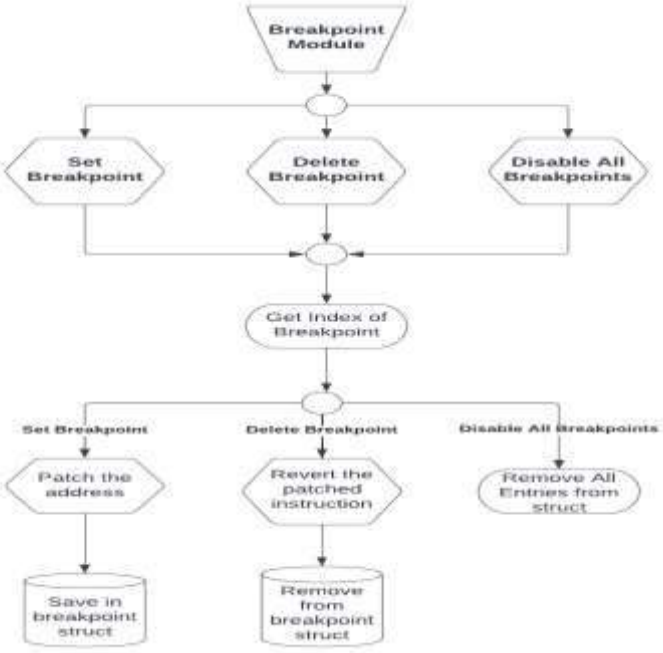
02. Break:

Fuzzy Dbg has a special structure dedicated to handling and saving breakpoint related information. To minimize p trace overhead, the implementation checks if the breakpoint is an already existing trace point, if so, it records the breakpoint without any redundant calls to p trace. All breakpoints otherwise set are patches made to the required instruction with the help of SIGTRAP interrupt via ptrace. From Fig.3, the implementational flow of the breakpoint module can be visualized.

03. Elf Info:

This module is for emitting the most basic information about an ELF. It parses the ELF file header, program headers, and sections. All static and dynamic symbols from the ELF as well as the standard library of the system are extracted by appropriate algorithms. Fig.5 depicts how the elf info module extracts all symbols from ELF and LIBC.


Attach Module
Fork
Child
Exec and Set Ptrace
Parent
Await for Events
Process Events
Detect Tracepoints, Breakpoints and any Signals
Process Tracepoints, Breakpoints and any other Signals


Breakpoint Module
Set Breakpoint
Delete Breakpoint
Disable All Breakpoints
Get Index of Breakpoint
Set Breakpoint
Delete Breakpoint
Disable All Breakpoints
Patch the address
Revert the patched instruction
Remove All Entries from struct
Save in breakpoint struct
Remove from breakpoint struct

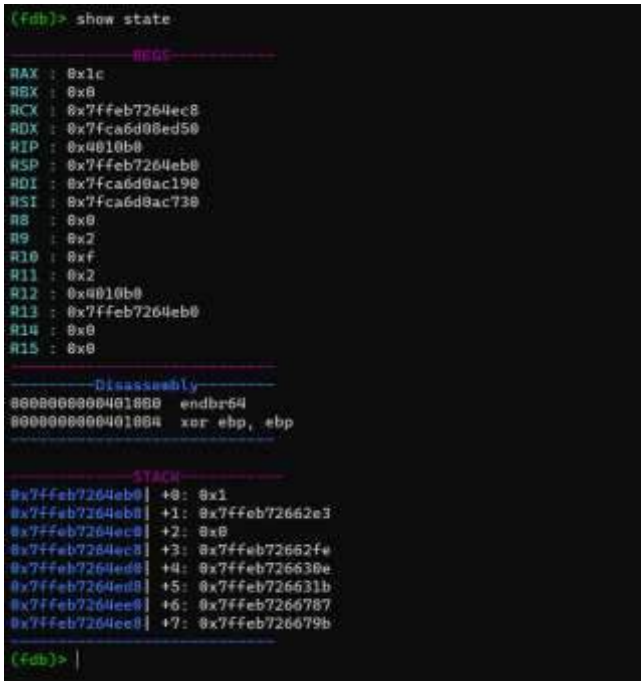
```
struct event {
enum {
            EVENT_UNKNOWN,
            EVENT_NONE,
            EVENT_SIGNAL,
            EVENT_EXIT,
            EVENT_EXIT_SIGNAL,
            EVENT_BREAKPOINT
   } type;
union {
        intretval; /* EVENT_EXIT */
        int signum; /*For other sigs*/
   } u;
};
```

Fig.4 Structure of an event

The show module is responsible for printing the requested specific states of the debugee at the current state with respect to the control flow. It displays the state of registers, disassembly, stack, virtual memory maps, all static and dynamic symbols as well as the execution trace as per desire of the user. From Fig.6 , we can see the state of registers, stack and disassembly at a given point of execution while debugging a binary.

04. Tracer Engine:

This module is responsible for creating trace points to create an execution trace for future analysis. Initially, the static and dynamic symbol tables are extracted from the binary as well as the shared library of the system. The shared library is parsed due to the fact that binary does not have any information about the addresses of the dynamic functions. After recording all the necessary static and dynamic functions along with their addresses, the module sets debugger interrupts at every function. When the particular function is hit, the function name is recorded in a list. From Fig.9 , we can see how the module has been implemented.



05. Fuzzer:

Fuzzy Dbg with its inbuilt fuzzer comes with the mutation engine written from scratch. The component uses polymorphic code to alter the payload at runtime to feed the debugee, generating a crash thereby and logging the crashing seed. From Fig.7, the implementation flow of the fuzzer can be observed.

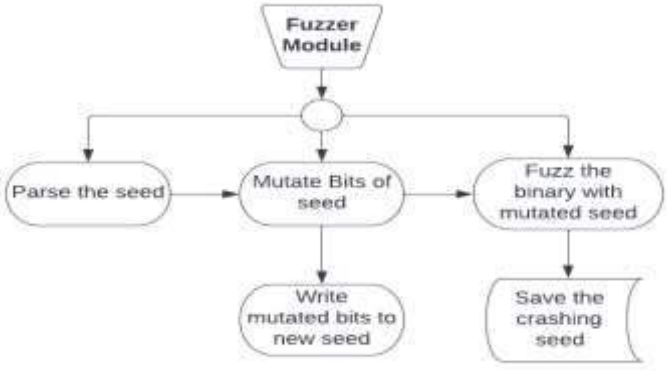


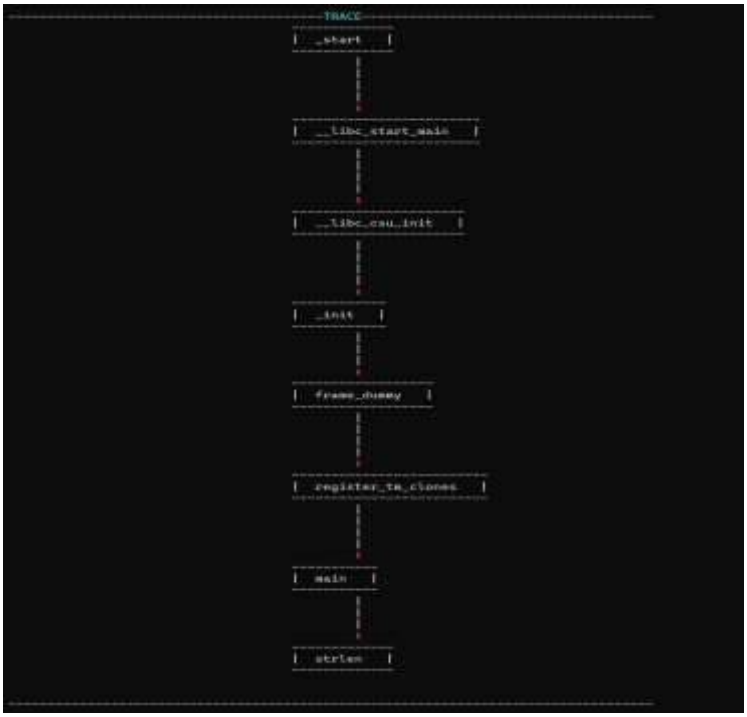
Fig.7 Implementational flow of fuzzer

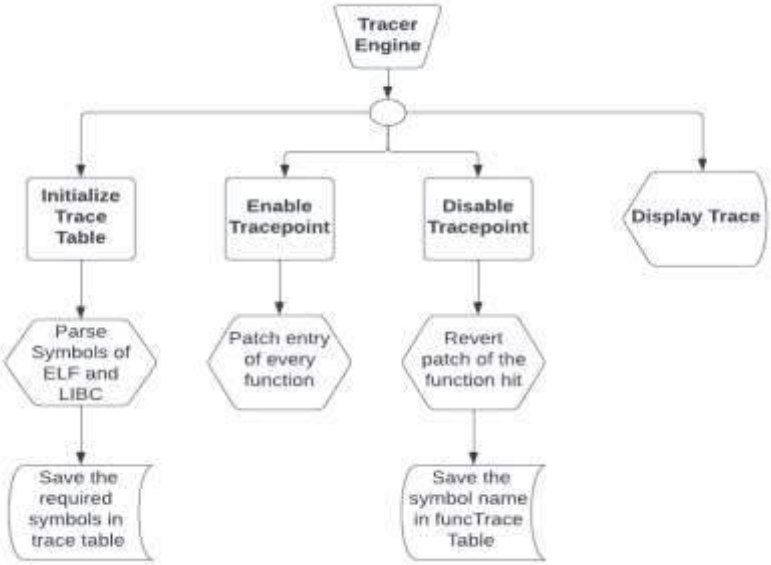
IV. RESULTS:

FuzzyDbg currently supports only x86-64 bit ELFs but can successfully attach to a process, single step through the assembly, set breakpoints at various segments of code, show various debug information like registers, assembly, stack ,virtual maps, runtime addresses of static and dynamic symbols, execution trace and fuzz the process on demand with a supplied seed all the while providing a simple interface to interact with it. fig.10 shows the fuzzer in action, the initial seed is mutated according to our mutation algorithm and feeded as an input to a binary which is prone to the buffer overflow vulnerability.

From fig.8 the practicality of the execution trace of a binary can be accessed. The executable used here for debugging segfaults at snprintf due to not having the correct argument type. The trace helps analyze all the functions which have been hit in the process of triggering the crash.

From fig.11, we can see a table showing the detection times of the fuzzer with various vulnerability classes - stack overflow, heap overflow, integer overflow, use after free. Here, we can see the effectiveness of having a






```
(fdb)> fuzz ./seed 20
[+] File length : 0x53
[>] Flipping and mutating up to 19 bytes.
[>] Fuzzing for 20 iterations...
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
[*] WIFEXITED: Exit Status: 0
ok
*** stack smashing detected ***: terminated
[>] Crashes: 1
```

Vulnerability Class	No of Mutations required for initial seed	Time (in sec) required for performing the mutations and detecting the vulnerability
Stack Buffer Overflow	30	0.036377
Heap Buffer Overflow	40	1.531245
Heap Use After Free	31	0.93123
Integer Overflow	23	0.68340

V. CONCLUSION:

The paper addresses the needs of a modern debugger: an inbuilt fuzzer to fuzz out critical vulnerabilities right inside the debugger and a detailed execution trace as and when a crash triggers or when the user desires. As part of future work, the debugger can be expanded to the realms of various other architectures and robust support for multi-threaded environments. To make it a truly complete debugging solution, GUI based control flow graphs to visualize basic blocks can also be designed within the debugger itself.

VI. ACKNOWLEDGEMENTS:

The authors would like to thank Asst. Prof. Vipin Pavithran, our project guide Asst. Prof. Jaya shree Narayanan and other anonymous reviewers for their valuable guidance and support. This work is supported by the Amrita Center for Cyber security Systems and Networks.

VII. REFERENCES:

1. Engin Durmaz, M. Borahan Tümer. Intelligent software debugging: A reinforcement learning
2. approach for detecting the shortest crashing scenarios.Expert Systems with Applications
3. Qin Zhao, Rodric Rabbah, Saman Amarasinghe, Larry Rudolph, Weng-Fai Wong. How to Do a Million Watch points: Efficient Debugging Using Dynamic Instrumentation. ICCSEE.
4. Chao Zhang, Lei Duan, Tao Wei, Wei Zou. SecGOT: Secure Global Offset Tables in ELF Executables. ICCSEE Conference.
5. Thomas Dupriez, Guillermo Polito, Stéphane Ducasse. Analysis and exploration for new generation debuggers. International Workshop on Smalltalk Technology IWST'17, Sep 2017, Maribor, Slovenia. pp.5:1–5:6, ff10. 1145/3139903. 3139910ff. ffhal-01585338f/file/Dup17 a – IWST - New Generation Debuggers.pdf
6. AmatCama, Yan Shoshitaishvili, Christopher Kruegel, and Giovanni Vigna. How the ELF Ruined Christmas. USENIX Security Symposium.
7. Tim Bird, Measuring Function Duration with Ftrace. Linux Symposium.
8. Sanjay Bhansali, Wen-Ke Chen, Stuart de Jong, Andrew Edwards, Ron Murray, Milenko Drini´c, Darek Miho˘cka, Joe Chau. Framework for Instruction-level Tracing and Analysis of Program Executions. Usenix Conference.

Automatic Pain Recognition Techniques: A State-of-the-art Review

Patil Manisha Shanataram
Department of Computer Engineering (Assistant Professor)
SSVPS B S D College of Engneering (DBATU Lonere)
Dhule, India patmanisha@gmail.com

Patil Hitendra Dhansing
Department of Computer Engineering (Professor)
SSVPS B S D College of Engneering (DBATU Lonere)
Dhule, India hitendradpatil@gmail.com

Abstract—Pain is a dynamic and subjective experience that can be difficult to measure. Automated clinical pain assessment method offers a lot of potential, and they're not widely employed in medical practice presently. There is now an a need for a comprehensive and precise method to identify acute pain among intensive care units in order to assist professionals in dispensing pain relievers at the proper dosage and on time. We review and discuss autonomous pain identification algorithms in this article also provide an introduction of pain processes and reactions, as well as a discussion of commonly used clinical pain assessment techniques and shared datasets.

Keywords—pain, recognition, feature, classification

I. INTRODUCTION

Pain is a dynamic occurrence that has been still not clearly realized. The standard statement of pain is "a stressful associated with emotions response generated by probable muscle injury, or characterized in relations of certain destruction"[48]. Basic research, on the other hand, continues to advance scientific knowledge of pain, and there is an active discussion about changing the definition [49-50]. Such kind of pain, known as acute pain, supports in the identification of potentially hazardous situations, the prevention of tissue damage, and the facilitation of recovery by preventing behaviours which may cause further tissue damage [50]. Many humans, as well as community overall,

are affected by pain. The growing demand for pain treatment has been aided by advances in medicine: Many people now a days are suffering diseases that were formerly deadly, such as HIV, cancer, and cardiovascular disease. However, they will have chronic pain as a result of either the existing disease or the surgery or even after the sickness has been cleared, by neurological damage caused by the condition [52]. Chemotherapy, Surgery, and radiotherapy are all common treatments that inflict pain [31]. Persistent pain has serious consequences for the person in pain, as well as her friends and family members. Scientifically valid pain assessment is needed for diagnostic process, selecting a suitable treatment, evaluating progress, and deciding whether such a treatment should be maintained or improved. So that, assessing and managing pain is important besides for providing relief. But also for eliminating together instantaneous and lasting implication as decreases life eminence however, it also completely undermines the neurological system [41]. Unrelieved pain can rise to chronic pain condition, which also is marked by restricted movement, weakened immunity, difficulty concentrating, obesity, and sleep disturbances. However, incorrect therapy might cause complications and complications for patients.

Although advances in technology and knowledge, pain is still mismanaged [52-54], [69]. However this is a prevalent difficult, it disproportionally disturbs patients with weak message skills, that are unable to communicate overall pain perception and those who's express has low ecological validity. In the next decade, autonomous pain assessment systems based on pain behaviours will be developed (video facial terminologies, movements of body and vocalizations) and Physiological responses would be used to enhance current pain diagnostic tools in order to achieve better pain treatment. In comparison to conventional assessment methods, it might regularly monitor pain. This could lead to improved treatment outcomes, such as besides facilitating early diagnosis for patients being unable to request assistance on their own. Moreover, automatic structures will be much more accurate rather than a person observer, whom assessment will be affected with different aspects like as the patient's appearance or relationship to the patient [53, 55, 69, 72, 73].

II. PAIN MECHANISMS AND RESPONSES

Pain is a unique, individual, sensory experience that originates in the brain. Pain is more than just a physical sensation [56]. The pain feeling should be identified from the cause of the pain (such as muscle destruction caused by nerve injury), the pain reaction (vocal communication and non- verbal Indicators) and pain assessment. The cause of pain is usually identifiable, and it can be managed by intentional pain stimulus.

A. Biological Mechanisms

The pain method involves numerous aspects of the neurological structure. The development usually starts with unpleasant mechanical, chemical cold, heat, or inflammatory stimuli activating sensory nerve cells. These signals stimulate nociceptors, which become primary sensory neurons with noxious stimuli-detecting specific receptors. The induced electrical impulses are transmitted to the spinal cord through nociceptive fibers. Excitatory network and also inhibitory interneuron network in the spinal cord can be stimulated, resulting in a protective reflexive retraction response. The perceptual exclusionary experience of pain is the resultant of next processing of nociceptive information in several spinal organs. Whereas a nociceptive signal usually causes pain, numerous factors can influence this response.

B. Biological Responses

Connections among neural network are complex, involved in Pain perceptions and autonomic control [57] because an increase in sympathetic outflow, leading in modifications in physiological signals that can be measurable [58].

Skin conductance is a signal [59] that changes in response to pain and is automatically modulated. Even though sweat organs are just stimulated by sympathetic excitatory sensory neurons [57], increased sympathetic flow in pain response leads sweat to be released into apertures on the surface of skin [59]. Heat influences the electrical performance of the skin (electrodermal, EDA and electrodermal activity), enhancing electrode potential until underarm sweat is evaporated or reabsorbed, sympathetic nervous system excitation also creates significant cardiovascular consequences. It has an influence on the heart

rate [60], this induces tachycardia and heart rate irregularity, an indicator of autonomic heart rate control. Especially, Pain massively increases different frequencies power, as evaluated by energy power spectrum. Moreover, Peripheral vascular resistance and stroke volume also are boosted by pain.

C. Behavioral Responses

Facial expressions, gestures, and speech patterns are characteristics of behavioural pain reactions. Chronic pain typically results in significant changes in daily behaviour and public communication. Presently, various pain-related facial terms that happen comparatively constantly across a wide variety of laboratory pain situations and evaluation pain methods [61]. Consequently, the intensity of facial movement's increases as the strength of painful stimuli increases [62]. The majority of pain-related bodily motions help to defend against future harm and to relieve unpleasantness. Pain behaviour also includes paralinguistic vocalisations (laughing, groaning, and moaning) and sound quality features including loudness, and insecurity detected throughout voiced report [61].

D. Emotion and Pain

Pain is classified as either a sensory or also an emotional reaction. It has an emotive element that covers a wide range of feelings, the majority of which are destructive and connected to the discomfort to potential consequences. Aggression and sorrow also performance key roles, particularly in chronic pain [62]. Lastly, physiological and behavioural responses indications and their groupings, the issue of whether pain can be consistently and objectively separated from its associated emotions remains unresolved.

III. TOOLS FOR PAIN ASSESSMENT IN CLINICAL APPLICATION

In clinical practice, pain is typically based mainly on the patient's report intensity and variables that relieve and increase the pain. Self-reporting is the explicit presentation of pain associated information through a person in pain,

generally through verbal or gestures such as directing to an illustration that signifies their feelings in answer to a request. Patient care regulations underline that report is the utmost reliable method of measuring pain if the person is able to speak [63-69].

IV. DATASETS

The best frequently used database is the BioVid Heat Pain Database [35]. It was acquired through a partnership between the University of Magdeburg's Neuro-Information groups. Total 90 patients were submitted to four intensities of experimentally induced heat sensitivity. Table 1 represents the properties of publicly accessible datasets for pain recognition research.

V. METHODOLOGIES OF PAIN RECOGNITION

We directed an efficient literature search, as described in introduction, to evaluate the existing automated pain recognition approaches. In the next subsections, we explore the recognition input systems and the system processing techniques.

A. Modalities and Sensors

Pain assessment needs at minimum single sensor information stream to share information to the machine. A medium of this type is sometimes referred to as a modality. The most significant pain recognition modalities may be classified into two parts: behaviour and physiology. Face expression; body motions like rubbing, guarding, and skull movements; vocalisations and vocal arguments that can be shared by communication and may contain self-report information are all examples of behavioural modalities. In the physiology area, brain action, cardiac action are all of interest. A uni-modal system consist of only one modality; a multimodal approach takes input from multiple modalities.

1) Camera-Based Methodology: To date, the large majority of pain detection systems have relied at camera imagery with facial expressions. The majority of

existing pain recognition research focused on Facial expression modality. In general, cameras have a narrow range of view, thus understanding images is much more complicated than processing other instrument inputs. Cameras, on the other hand, are non-contact devices that may be more comfortable or patients and more efficient for healthcare workers thancontact-based sensors [42].

2) Contact-Sensor Techniques: The direct interaction sensors EDA and ECG have been the second-most extensively utilised. Since, the introduction of the BioVid database, followed by sEMG of the trapezius muscle-Sensor techniques. Pain was identified using a variety of physiological indicators acquired from electronic flow spreadsheets in institutions. EDA consistently outperforms the only one modalities that have evaluated [25], [33].

TABLE I. Databases for Pain Recognition that are available to the public for research

Database	Subjects	Stimuli	Data Modalities (D)
BioVid Heat Pain[30], [66], [67]	Total 90 adult participants	14k heat pain (20 repetitions 4 intensities2parts 90 adult patient)	Video signal, biomedical indicators
BP4D-Spontaneous [64]	Total 41 adults	emotion elicitation, 41 cold pressor task	Color and 3D video of face
BP4D+[68]	Total 140 adults	emotion elicitation ,140 cold pressor task	Face of color video 3D and thermal,medical signal such as respiration rate, blood pressure, heart rate andEDA

UNBC McMaster Shoulder Pain [65]	Total 25 patients shoulder pain	Total 200 video signal,	Face video such as low resolution, social interaction and talking
MIntPAIN [41]	Total 20 adults	40 stimuli in 4 intensities of 2 trials and 20 participant of 2k electrical pain	Fave video of color, thermal and depth

TABLE II. Bio-Vid Database has been used to evaluate Pain Recognition systems

Author	Feature	Classification Model
M. Amirian et al.[71]	Descriptor with time series statics signal.	Radial Basis Function networks,
S. Gruss et al.[72]	159 features based on statistical analyses	radial basis function kernel and SVM
M. Kachele et al. [22]	Head pose, peak height and difference, B.W. and entropy.	SVM and Random forest
M. Kachele et al. [23]	Physiological signal and geometric-based and appearance-based from video.	Random Forest (RF)
M. Kachele et al.[24]	Mean and the standard deviation from EMG.	KNN and Random forest
M. Kachele et al. [25]	Frame level, skewness, spectral entropy,	Random Forest
Lopez-M. 17 [26]	Calculated the Skin conductance (SC) and Electrocardiogram (ECG) features	multi-task neural network (MT-NN)

Lopez-M.18 [27]	Mean, max, range, AUC	Recurrent neural network-based regression algorithm
Walter 14 [28]	Shannon entropy, and heart rate	Decision Trees, K-NN, andSVMs.
Walter 15 [29]	Features based on signal amplitude and frequency also based on entropy, stationary and statistical moments.statistical parameters from video signal	KNN and Random forest
Werner 14 [31]	Features of amplitude from (GSR) ,frequency features from (EMG) at trapezius muscle and three heart rate variability features from (ECG)	Random Forest
Werner 17 [32]	Measure facial activity and expressiveness	Random Forest
Yang 16 [18]	Texture information form face	Support Vector Machine(SVM)
F. Pouromran [70]	22 features capture properties such as linear and nonlinear autocorrelation, successive differences.	Support Vector Regressor, XGBoost,KNN, NeuralNetworks

B. Features and Models

The raw data is analysed for patterns that can be used to predict a pain state in the person being studied. Features are collected from the input signal, which are a more exclusionary and generally lesser dimensional representation. There are an amount of diverse kinds of features classified as (1) learned

features, (2) generic features, (3) hand-designed. Generic features are useful in other domains, but pain recognition they not provide better result. Such as, LBP image features. Hand- crafted features are constructed for a specific objective through dedicated knowledge; they are generally simple to interpret and low-dimensional. Another feature such as use during the training procedure, the extracted features is optimized for the specific task termed as learned features. These are typically higher dimensionality and hard to interpret.

1)Face Expression Features: Feature extraction is a process in the processing system during camera-based facial pain expression analysis. It might include 1) identifying facial characteristics (points around the mouth, brows and eyes and among others) 2) To improvement invariance to translation, scaling, and rotation, utilize facial texture. A range of frame- based features have utilized to pain recognise [1], [8], [11], [15-16], [19], [20], [35]; simplest pixel representations represent illustrations using generic appearance features [3], [21], [37], LBP [8], [12], [16], [18], [20], [35], HOG [1], [6], [12], [34], Discrete Cosine Transform (DCT) [35], Gabor [9], [12], [20], [39], [44], Scale Invariant Feature Transform (SIFT) [10], [14], neural network-learned features [7], [13], [17], [21], [41].

2) Physiological Features: All other sensor signals, apart from camera images and neuroimaging, are interpreted as time signal. In the evaluation of sEMG, EDA signals, we identify numerous variations of Time series signal Descriptors (TSD). Walter [48] investigated the effectiveness of amplitude, variability, and frequency. sEMG (corrugator, zygomaticus muscle and trapezius) and stationarity, EDA frequency, entropy, linearity and, entropy features for pain recognition. Based on physiological objectives, EDA has divided into two aspects while retrieving information from both components independently [23], [25], [28], [33].

3) Recognition Models: After feature extraction, the framework that relates the features to the implicit pain status is the second essential processing component. The model may also include data fusion, especially for system integration in a multimodal system, which can be done at the decision, feature, or intermediate levels. The most of methods utilise Support Vector Machines

(SVMs) to categorise pain, either linearly or with a Radial Basis Function (RBF) kernel [11], [28], [30], [38], [39], [42]. Relevance Vector Regression generates continuous-valued output [1], [5].and related Support Vector Regression [2], [4], [16], [20] models. Random Forests are another popular model (RF) [29], [32], [35, [36], [41] variations of Conditional Random Fields (CRF) [3], [8], [15], Nearest Neighbor (NN) classifiers [6], [19], [24], [36], [37], and various neural networks. Convolutional Neural Networks are one of the CNN architectures utilised for pain identification [13], [17], [21], [41], Radial Basis Function (RBF) networks [25], Long Short-Term Memory (LSTM) networks [8], [13], [41], Data fusion is a process of integrating various modalities, features, judgement score, or even other information sources to create a single final prediction [22], [23], [25], [33].

C. Ground Truth

Different objectives being explored by the evaluated automated pain assessment systems. The most common are evaluating the existence of pain (a binary classification) and assessing the intensity of pain. Such systems require adequate ground truth for development and evaluation. The majority of other study employed the provided stimulus as ground truth, either with [46], [47] or instead of customized assessment [39], [40], [43]. The most of these studies aim to investigate the occurrence of pain or anticipate the intensity of symptoms in separate categories [40],[41], [45], The later researches revealed that an integrated computer vision system outperformed skilled human observers in identifying realistic from manufactured pain signals on the face.

VI. CONCLUSION

Health services are complicated processes which include interaction among people, organizations, and equipment. Researcher kept in mind these information as an initial phase to design an intelligent health-care system. This article provides a complete overview of pain assessment techniques that depend on several machine learning techniques. The dada set utilized for pain assessment was consider as video or image based facial expression or biomedical signal.

REFERENCES

1. Egede, Joy, Michel Valstar, and Brais Martinez. "Fusing Deep Learned and Hand-Crafted Features of Appearance, Shape, and Dynamics for Automatic Pain Estimation." 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017. https://doi.org/10.1109/fg.2017.87
2. Florea, Corneliu, Laura Florea, Raluca Butnaru, Alessandra Bandrabur, and Constantin Vertan. "Pain Intensity Estimation by a Self-Taught Selection of Histograms of Topographical Features." Image and Vision Computing 56 (2016): 13–27. https://doi.org/10.1016/j.imavis.2016.08.014.
3. Ghasemi, Afsane, Xinyu Wei, Patrick Lucey, Sridha Sridharan, and Clinton Fookes. "Social Signal Processing for Pain Monitoring Using a Hidden Conditional Random Field." 2014 IEEE Workshop on Statistical Signal Processing (SSP), 2014. https://doi.org/10.1109/ssp.2014.6884575.
4. Hong, Xiaopeng, Guoying Zhao, Stefanos Zafeiriou, Maja Pantic, and Matti Pietikäinen. "Capturing Correlations of Local Features for Image Representation." Neurocomputing 184 (2016): 99–106. https://doi.org/10.1016/j.neucom.2015.07.134.
5. Kaltwang, Sebastian, Sinisa Todorovic, and Maja Pantic. "Doubly Sparse Relevance Vector Machine for Continuous Facial Behavior Estimation." IEEE Transactions on Pattern Analysis and Machine Intelligence 38, no. 9 (2016): 1748–61. https://doi.org/10.1109/tpami.2015.2501824.
6. Khan, Rizwan Ahmed, Alexandre Meyer, Hubert Konik, and Saida Bouakaz. "Pain Detection through Shape and Appearance Features." 2013 IEEE International Conference on Multimedia and Expo (ICME), 2013. https://doi.org/10.1109/icme.2013.6607608.
7. Kharghanian, Reza, Ali Peiravi, and Farshad Moradi. "Pain Detection from Facial Images Using Unsupervised Feature Learning Approach." 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016. https://doi.org/10.1109/embc.2016.7590729.
8. Martinez, Daniel Lopez, Ognjen Rudovic, and Rosalind Picard. "Personalized Automatic Estimation of Self- Reported Pain Intensity from

Facial Expressions." 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017. https://doi.org/10.1109/cvprw.2017.286.

9. Lo Presti, Liliana, and Marco La Cascia. "Boosting Hankel Matrices for Face Emotion Recognition and Pain Detection." Computer Vision and Image Understanding 156(2017):19–33. https://doi.org/10.1016/j.cviu.2016.10.007.
10. Neshov, Nikolay, and Agata Manolova. "Pain Detection from Facial Characteristics Using Supervised Descent Method." 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2015. https://doi.org/10.1109/idaacs.2015.7340738.
11. Rathee, Neeru, and Dinesh Ganotra. "A Novel Approach for Pain Intensity Detection Based on Facial Feature Deformations." Journal of Visual Communication and Image Representation 33 (2015): 247–54. https://doi.org/10.1016/j.jvcir.2015.09.007.
12. Rathee, Neeru, and Dinesh Ganotra. "Multiview Distance Metric Learning on Facial Feature Descriptors for Automatic Pain Intensity Detection." Computer Vision and Image Understanding 147 (2016): 77–86. https://doi.org/10.1016/j.cviu.2015.12.004.
13. Rodriguez, Pau, Guillem Cucurull, Jordi Gonalez, Josep M. Gonfaus, Kamal Nasrollahi, Thomas B. Moeslund, and F. Xavier Roca. "Deep Pain: Exploiting Long Short- Term Memory Networks for Facial Expression Classification." IEEE Transactions on Cybernetics, 2017, 1–11. https://doi.org/10.1109/tcyb.2017.2662199.
14. Rudovic, Ognjen, Vladimir Pavlovic, and Maja Pantic. "Context-Sensitive Dynamic Ordinal Regression for Intensity Estimation of Facial Action Units." IEEE Transactions on Pattern Analysis and Machine Intelligence 37, no. 5 (2015): 944–58. https://doi.org/10.1109/tpami.2014.2356192.
15. Ruiz, Adria, Xavier Binefa, and Joost Van de Weijer. "Regularized Multi-Concept Mil for Weakly-Supervised Facial Behavior Categorization."

Proceedings of the British Machine Vision Conference 2014, 2014. https://doi.org/10.5244/c.28.13.

16. Rupenga, Moses, and Hima B. Vadapalli. "Automatic Spontaneous Pain Recognition Using Supervised Classification Learning Algorithms." 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016. https://doi.org/10.1109/robomech.2016.7813150.
17. Wang, Feng, Xiang Xiang, Chang Liu, Trac D. Tran, Austin Reiter, Gregory D. Hager, Harry Quon, Jian Cheng, and Alan L. Yuille. "Regularizing Face Verification Nets for Pain Intensity Regression." 2017 IEEE International Conference on Image Processing (ICIP),2017. https://doi.org/10.1109/icip.2017.8296449.
18. Yang, Ruijing, Shujun Tong, Miguel Bordallo, Elhocine Boutellaa, Jinye Peng, Xiaoyi Feng, and Abdenour Hadid. "On Pain Assessment from Facial Videos Using Spatio-Temporal Local Descriptors." 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016.https://doi.org/10.1109/ipta.2016.7820930.
19. Zafar, Zuhair, and Nadeem Ahmad Khan. "Pain Intensity Evaluation through Facial Action Units." 2014 22nd International Conference on Pattern Recognition, 2014.https://doi.org/10.1109/icpr.2014.803.
20. Zhao, Rui, Quan Gan, Shangfei Wang, and Qiang Ji. "Facial Expression Intensity Estimation Using Ordinal Information." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.https://doi.org/10.1109/cvpr.2016.377.
21. Zhou, Jing, Xiaopeng Hong, Fei Su, and Guoying Zhao. "Recurrent Convolutional Neural Network Regression for Continuous Pain Intensity Estimation in Video." 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016. https://doi.org/10.1109/cvprw.2016.191.

22. Kächele, Markus, Philipp Werner, Ayoub Al-Hamadi, Günther Palm, Steffen Walter, and Friedhelm Schwenker. "Bio-Visual Fusion for Person-Independent Recognition of Pain Intensity." Multiple Classifier Systems, 2015, 220–30. https://doi.org/10.1007/978-3- 319-20248-8_19.
23. Kächele, Markus, Patrick Thiam, Mohammadreza Amirian, Philipp Werner, Steffen Walter, Friedhelm Schwenker, and Günther Palm. "Multimodal Data Fusion for Person-Independent, Continuous Estimation of Pain Intensity." Engineering Applications of Neural Networks, 2015, 275–85. https://doi.org/10.1007/978-3-319-23983-5_26.
24. Kachele, Markus, Patrick Thiam, Mohammadreza Amirian, Friedhelm Schwenker, and Gunther Palm. "Methods for Person-Centered Continuous Pain Intensity Assessment from Bio-Physiological Channels." IEEE Journal of Selected Topics in Signal Processing 10, no. 5 (2016): 854–64.https://doi.org/10.1109/jstsp.2016.2535962.
25. Kächele, Markus, Mohammadreza Amirian, Patrick Thiam, Philipp Werner, Steffen Walter, Günther Palm, and Friedhelm Schwenker. "Adaptive Confidence Learning for the Personalization of Pain Intensity Estimation Systems." Evolving Systems 8, no. 1 (2016): 71–83. https://doi.org/10.1007/s12530-016-9158-4.
26. Lopez-Martinez, Daniel, Ognjen Rudovic, and Rosalind Picard. "Physiological and Behavioral Profiling for Nociceptive Pain Estimation Using Personalized Multitask Learning." arXiv.org, November 10, 2017. https://arxiv.org/abs/1711.04036.
27. Lopez-Martinez, Daniel, and Rosalind Picard. "Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018. https://doi.org/10.1109/embc.2018.8513575.
28. Walter, Steffen, Sascha Gruss, Kerstin Limbrecht- Ecklundt, Harald C.

Traue, Philipp Werner, Ayoub Al- Hamadi, Nicolai Diniz, Gustavo Moreira da Silva, and Adriano O. Andrade. "Automatic Pain Quantification Using Autonomic Parameters." Psychology &Neuroscience 7, no. 3 (2014): 363–80. https://doi.org/10.3922/j.psns.2014.041.

29. Walter, Steffen, Sascha Gruss, Markus Kächele, Friedhelm Schwenker, philipp Werner, Ayoub Al- Hamadi, Adriano Andrade, Gustavo Moreira, and Harald Traue. "Data Fusion for Automated Pain Recognition." Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare, 2015.https://doi.org/10.4108/icst.pervasivehealth.2015. 259166.
30. Werner, Philipp, Ayoub Al-Hamadi, Robert Niese, Steffen Walter, Sascha Gruss, and Harald Traue. "Towards Pain Monitoring: Facial Expression, Head Pose, a New Database, an Automatic System and Remaining Challenges." Procedings of the British Machine Vision Conference 2013, 2013. https://doi.org/10.5244/c.27.119.
31. Werner, Philipp, Ayoub Al-Hamadi, Robert Niese, Steffen Walter, Sascha Gruss, and Harald C. Traue."Automatic Pain Recognition from Video and Biomedical Signals." 2014 22nd International Conference on Pattern Recognition, 2014. https://doi.org/10.1109/icpr.2014.784.
32. Werner, Philipp, Ayoub Al-Hamadi, and Steffen Walter."Analysis of Facial Expressiveness during Experimentally Induced Heat Pain." 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017. https://doi.org/10.1109/aciiw.2017.8272610.
33. Thiam, Patrick, and Friedhelm Schwenker. "Multi- Modal Data Fusion for Pain Intensity Assessment and Classification." 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), 2017. https://doi.org/10.1109/ipta.2017.8310115.
34. Thiam, Patrick, Viktor Kessler and Friedhelm Schwenker. "Hierarchical Combination of Video Features for Personalised Pain Level Recognition." ESANN (2017).

35. Aung, Min S., Sebastian Kaltwang, Bernardino Romera- Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, et al. "The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal Emopain Dataset." IEEE Transactions on Affective Computing 7, no. 4 (2016): 435–51. https://doi.org/10.1109/taffc.2015.2462830.
36. Zamzmi, Ghada, Chih-Yun Pai, Dmitry Goldgof, Rangachar Kasturi, Yu Sun, and Terri Ashmeade. "Automated Pain Assessment in Neonates." Image Analysis, 2017, 350–61. https://doi.org/10.1007/978-3- 319-59129-2_30.
37. Adibuzzaman, Mohammad, Colin Ostberg, Sheikh Ahamed, Richard Povinelli, Bhagwant Sindhu, Richard Love, Ferdaus Kawsar, and Golam Mushih Ahsan. "Assessment of Pain Using Facial Pictures Taken with a Smartphone." 2015 IEEE 39th Annual Computer Software and Applications Conference, 2015.https://doi.org/10.1109/compsac.2015.150.
38. Ashouri, Sajad, Mohsen Abedi, Masoud Abdollahi, Farideh Dehghan Manshadi, Mohamad Parnianpour, and Kinda Khalaf. "A Novel Approach to Spinal 3-D Kinematic Assessment Using Inertial Sensors: Towards Effective Quantitative Evaluation of Low Back Pain in Clinical Settings." Computers in Biology and Medicine 89 (2017): 144–49. https://doi.org/10.1016/j.compbiomed.2017.08.002.
39. Bartlett, Marian Stewart, Gwen C. Littlewort, Mark G. Frank, and Kang Lee. "Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions." Current Biology 24, no. 7 (2014): 738–43. https://doi.org/10.1016/j.cub.2014.02.009.
40. Chu†, Yaqi, Xingang Zhao, Jun Yao†, Yiwen Zhao, and Zhenwei Wu. "Physiological Signals Based Quantitative Evaluation Method of the Pain." IFAC Proceedings Volumes 47, no. 3 (2014): 2981–86.

https://doi.org/10.3182/20140824-6-za-1003.01420.

41. Haque, Mohammad A., Ruben B. Bautista, Fatemeh Noroozi, Kaustubh Kulkarni, Christian B. Laursen, Ramin Irani, Marco Bellantonio, et al. "Deep Multimodal Pain Recognition: A Database and Comparison of Spatio-Temporal Visual Modalities." 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018. https://doi.org/10.1109/fg.2018.00044.
42. Oshrat, Yaniv, Ayala Bloch, Anat Lerner, Azaria Cohen, Mireille Avigal, and Gabi Zeilig. "Speech Prosody as a Biosignal for Physical Pain Detection." Speech Prosody 2016, 2016. https://doi.org/10.21437/speechprosody.2016-86.
43. Pourshoghi, Ahmad, Issa Zakeri, and Kambiz Pourrezaei. "Application of Functional Data Analysis in Classification and Clustering of Functional near-Infrared Spectroscopy Signal in Response to Noxious Stimuli." Journal of Biomedical Optics 21, no. 10 (2016): 101411. https://doi.org/10.1117/1.jbo.21.10.101411.
44. Susam, Busra T., Murat Akcakaya, Hooman Nezamfar, Damaris Diaz, Xiaojing Xu, Virginia R. de Sa, Kenneth Craig, Jeannie S. Huang, and Matthew S. Goodwin. "Automated Pain Assessment Using Electrodermal Activity Data and Machine Learning." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018. https://doi.org/10.1109/embc.2018.8512389.
45. Vatankhah, Maryam, and Amir Toliyat. "Pain Level Measurement Using Discrete Wavelet Transform." International Journal of Engineering and Technology 8, no. 5 (2016): 380–84. https://doi.org/10.7763/ijet.2016.v8.917.
46. Wang, Shuang, Xiaoqian Jiang, Zhanglong Ji, Robert El-Kareh, Jeeyae Choi, and Hyeoneui Kim. "When You Can't Tell When It Hurts: A Preliminary Algorithm to Assess Pain in Patients Who Can't Communicate."AMIA ...

Annual Symposium proceedings. AMIA Symposium. American Medical Informatics Association, November 16, 2013. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC39001 56/.

47. Werner, Philipp, Ayoub al-Hamadi, and Robert Niese. "Comparative Learning Applied to Intensity Rating of Facial Expressions of Pain." International Journal of Pattern Recognition and Artificial Intelligence 28, no. 05 (2014):1451008.https://doi.org/10.1142/s0218001414510082.
48. Williams, Amanda C., and Kenneth D. Craig. "Updating the Definition of Pain." Pain 157, no. 11 (2016): 2420– 23. https://doi.org/10.1097/j.pain.0000000000000613.
49. Aydede, Murat. "Defending the IASP Definition of Pain." The Monist 100, no. 4 (2017): 439–64. https://doi.org/10.1093/monist/onx021.
50. AC, Williams. "Facial Expression of Pain: An Evolutionary Account." The Behavioral and brain sciences. 25,no.4(2002):439-55..doi: 10.1017/s0140525x02000080. PMID: 12879700.
51. Mäntyselkä, Pekka, Esko Kumpusalo, Riitta Ahonen, Anne Kumpusalo, Jussi Kauhanen, Heimo Viinamäki, Pirjo Halonen, and Jorma Takala. "Pain as a Reason to Visit the Doctor: A Study in Finnish Primary Health Care." Pain 89, no. 2 (2001): 175–80. https://doi.org/10.1016/s0304-3959(00)00361-4.
52. Turk, D. C., & Melzack, R. (2011). The measurement of pain and the assessment of people experiencing pain. In C. Turk & R. Melzack (Eds.), Handbook of pain assessment (pp. 3–16). The Guilford Press
53. Beck SL;Dunton N;Berry PH;Brant JM;Guo JW;Potter C;Spornitz B;Eaton J;Wong B; "Dissemination and Implementation of Patient-Centered Indicators of Pain Care Quality and Outcomes." Medical care. U.S. National Library of Medicine, 2018. https://pubmed.ncbi.nlm.nih.gov/30570589/.
54. Herr, Keela, Patrick J. Coyne, Margo McCaffery, Renee Manworren, and

Sandra Merkel. "Pain Assessment in the Patient Unable to Self-Report: Position Statement with Clinical Practice Recommendations." Pain Management Nursing 12, no. 4 (2011): 230–50. https://doi.org/10.1016/j.pmn.2011.10.002.

55. Zamzmi, Ghada, Rangachar Kasturi, Dmitry Goldgof, Ruicong Zhi, Terri Ashmeade, and Yu Sun. "A Review of Automated Pain Assessment in Infants: Features, Classification Tasks, and Databases." IEEE Reviews in Biomedical Engineering 11 (2018): 77–96. https://doi.org/10.1109/rbme.2017.2777907. Prkachin, Kenneth M., and Kenneth D. Craig. "Expressing Pain: The Communication and Interpretation of Facial Pain Signals." Journal of Nonverbal Behavior 19, no. 4 (1995): 191–205. https://doi.org/10.1007/bf02173080.
56. Stewart, Gordon, and Ajit Panickar. "Role of the Sympathetic Nervous System in Pain." Anaesthesia & Intensive Care Medicine 14, no. 12 (2013): 524–27.https://doi.org/10.1016/j.mpaic.2013.09.003.
57. Boucsein, Wolfram. "Electrodermal Activity." SpringerScience & Business Media, 2012. https://doi.org/10.1007/978-1-4614-1126-0.
58. Benedek, Mathias, and Christian Kaernbach. "Decomposition of Skin Conductance Data by Means of Nonnegative Deconvolution." Psychophysiology 47, no. 647-658 (2010). https://doi.org/10.1111/j.1469-8986.2009.00972.x.
59. Appelhans, Bradley, and Linda Luecken. "Heart Rate Variability and Pain: Associations of Two Interrelated Homeostatic Processes." Biological Psychology 77, no. 2 (2008):174–82. https://doi.org/10.1016/j.biopsycho.2007.10.004.
60. Prkachin, Kenneth M., and Patricia E. Solomon. "The Structure, Reliability and Validity of Pain Expression: Evidence from Patients with Shoulder Pain." Pain 139, no.2 (2008): 267–74. https://doi.org/10.1016/j.pain.2008.04.010.
61. Simon, Daniela, Kenneth D. Craig, Frederic Gosselin, Pascal Belin, and Pierre Rainville. "Recognition and Discrimination of Prototypical Dynamic Expressions of Pain and Emotions." Pain 135, no. 1 (2008): 55–64.

https://doi.org/10.1016/j.pain.2007.05.008.

62. Lumley, Mark A., Jay L. Cohen, George S. Borszcz, Annmarie Cano, Alison M. Radcliffe, Laura S. Porter, Howard Schubiner, and Francis J. Keefe. "Pain and Emotion: A Biopsychosocial Review of Recent Research." Journal of Clinical Psychology 67, no. 9 (2011): 942–68. https://doi.org/10.1002/jclp.20816.
63. Puntillo, Kathleen A., Ann B. Morris, Carol L. Thompson, Julie Stanik-Hutt, Cheri A. White, and Lorie
64. R. Wild. "Pain Behaviors Observed during Six Common Procedures: Results from Thunder Project II*." Critical Care Medicine 32, no. 2 (2004): 421–27.https://doi.org/10.1097/01.ccm.0000108875 35298.d2.
65. Kunz, M., and S. Lautenbacher. "The Faces of Pain: A Cluster Analysis of Individual Differences in Facial Activity Patterns of Pain." European Journal of Pain 18, no. 6 (2013): 813–23. https://doi.org/10.1002/j.1532-2149.2013.00421.x.
66. Zhang, Lin, Steffen Walter, Xueyao Ma, Philipp Werner, Ayoub Al-Hamadi, Harald C. Traue, and Sascha Gruss. "'BioVid Emo DB': A Multimodal Database for Emotion Analyses Validated by Subjective Ratings." 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016. https://doi.org/10.1109/ssci.2016.7849931.
67. Zhang, Zheng, Jeffrey M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, et al. "Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. https://doi.org/10.1109/cvpr.2016.374,
68. Brahnam, Sheryl, Chao-Fa Chuang, Frank Y. Shih, and Melinda R. Slack. "SVM Classification of Neonatal Facial Images of Pain." Fuzzy Logic and Applications, 2006, 121–28. https://doi.org/10.1007/11676935_15,
69. Gruss, Sascha, Mattis Geiger, Philipp Werner, Oliver Wilhelm, Harald C.

Traue, Ayoub Al-Hamadi, and Steffen Walter. "Multi-Modal Signals for Analyzing Pain Responses to Thermal and Electrical Stimuli." Journal of Visualized Experiments, no. 146 (2019). https://doi.org/10.3791/59057.

70. Pouromran, Fatemeh, Srinivasan Radhakrishnan, and Sagar Kamarthi. "Exploration of Physiological Sensors, Features, and Machine Learning Models for Pain Intensity Estimation." PLOS ONE 16, no. 7 (2021). https://doi.org/10.1371/journal.pone.0254108.
71. Amirian, Mohammadreza, Markus Kächele, and Friedhelm Schwenker. "Using Radial Basis Function Neural Networks for Continuous and Discrete Pain Estimation from Bio-Physiological Signals." Artificial Neural Networks in Pattern Recognition, 2016, 269–84. https://doi.org/10.1007/978-3-319-46182-3_23.
72. Gruss, Sascha, Roi Treister, Philipp Werner, Harald C. Traue, Stephen Crawcour, Adriano Andrade, and Steffen Walter. "Pain Intensity Recognition Rates via Biopotential Feature Patterns with Support Vector Machines." PLOS ONE 10, no. 10 (2015). https://doi.org/10.1371/journal.pone.0140330.
73. S. Jafar Ali Ibrahim and Thangamani, M "Proliferators and Inhibitors Of Hepatocellular Carcinoma", International Journal of Pure and Applied Mathematics (IJPAM) Special Issue of Mathematical Modelling of Engineering Problems Vol 119 Issue.15.July 2018
74. Compound feature generation and boosting model for cancer gene classification Ibrahim, S. Jafar Ali Ibrahim., Affir, A.M., Thangamani, M. International Journal of Engineering Trends and Technology, 2020, 68(10), pp. 48–51.

Iot Based Industrial Automation

[1]Dr.S.Kannappan, [a]K.kanchana, [b]B.Haritha, [c]K.Bhavana, [d]T.MadhuPriya, [e]G.Lokesh

[1]Associate Professor, ECE, GATES Institute of Technology, Gooty,
A B. Tech Students, ECE, GATES Institute of Technology, Gooty, AP

ABSTRACT:

The number of industries in a certain region is increasing year by year due to the increasing needs of people, and therefore intelligent industrialization automation can be used. Here the air quality in the industry and thenumber of products manufactured in the industry are monitored using gas detectors (to detect toxic or toxic gases) and IR sensors (to count the number of products manufactured) where there is no need for labor. Also to see estimated time of arrival of delivery vehicle GPS is used to getlatitude, longitude and estimated time of delivery vehicle in real time. The branch of the delivery vehicle is displayed via the APP thatcan bebuiltwith AndroidStudio,thisGPSis integrated in the NODE MCU.In thecaseofindustrialfireaccidents,thiscan be detected using the industry existing flame sensors built into the NODE MCU and the LCD screen where sprinklers are later activated.

INTRODUCTION:

Due to population growth, people's needs are increasing year by year, which leads to the development and establishment of more and more industries. In the past, the purpose of automation was to increase productivity and reduce the costs associated with human operators. When the word industry comes up, words like big, big and huge come to mind. Where monitoring the whole industry in retail handling is not an easy task, it requires many workers to check the function of the device and the working environment of the industry and track the package. Therefore, this experimental setup is proposed to reduce the human effort.

Industrial automation is the process of making industrial production processes more flexible and simple with greater efficiency. The integration of automation into industry leads to intelligent manufacturing solutions with improved product quality and productivity with reduced downtime and scrap. Industrial automation includes the use of various control devices such as PC/PLC, various sensor sand actuators, communication buses/modules, machine drives,HMI (Human Machine Interface) systems and other control devices.

Industrial automation:

This type of automation is most commonly used in automotive, computing and electronics, medical, telecommunications, consumer goods, and other industrial applications. The automation systems can be fixed, programmed, flexible and integrated systems.

Some of the types of industrial automation are listed below:

Numerically controlled machines:

These machines are computer-controlled machines that use computers to perform the control operations by sensing, processing, calculating and controlling the process variables. This automation is a programmed version of machine tools and is also knownasa CNC (Computerized Numerical Controlled) machine. These CNC machines are used incutting and milling applications for high accuracy and precise precision operation.

Computer Aided Manufacturing (CAM):

The entire manufacturing process (including production, planning and control) is automated through the use of numerically controlled machines, industrial robots and other types of automation devices. These automation systems also use computers to plan, design and shape the various products. Examples of this automation system are Computer Aided Design (CAD), Computer Aided Design and Drafting (CADD), and Computer Aided Process Planning (CAPP).

Industrialrobots:

These are a type of automated machines or devices that can perform the various tasks for a long period of time. These are mostly used in areas that are highly dangerous or hazardous to humans.

Now that we've seen a little about the layout of a typical industrial automation system, let's move on to discussing the different types of industrial automation systems.

LITERATUE SURVEY:-

On November 2014 the author Li Da Zu Proposed Internet of Things in industries [1]. Industrial Automation Using Internet of Things (IOT)In thispaper,they are developing a system which willautomatically monitor the industrialapplications and generate Alerts/Alarms or take intelligent decisions using concept of IoT. On 2014 the author Sadeque Reza Khan proposed GUI based Industrial Monitoring and control system [2]. IOT is achieved by using local networking standards and remotely controlling and monitoring industrial device parameters by using Raspberry Pi and Embedded web server Technology. Raspberry Pi module consists of ARM11 processor and Real Time Operating system whereas embedded web server technology is the combination of embedded device and Internet technology. Using embedded web server along with raspberry pi it is possible to monitor andcontrol industrial devices remotely by using local internet browser. The author Ayman Sleman & Reinhard Moeller proposed integrationof wirelesssen so rne twork service in too ther home industrial networks[3].They have developed new technologies that have allowed us to move from the First generation of the Internet into the current transition into the Fourth generation. This generation has been propelled by the concept of the Internet of Things (IoT). On 2013 the author Rajeev Piyare & Seong Ro Lee proposed Smart home control and Monitoring system using smart phone proposed "IOT BASED AUTOMATED TEMPERATURE AND HUMIDITY MONITORING AND CONTROL" [4]. A raspberry pi running with Linux OS coded with C++ program that retrieves the temperature as well as humidity readings and these values are sensed and sent to the internet. On 2011 the author Jinsoo

Han, Chang-Sic Choi, Wan-Ki park, Iiwoo Lee Green Home Energy Management System through comparison of energy usage between the same kinds of home appliances [5].

The concept of the internet of things was introduced by themembers of the radio frequency identification development community in 1999.This concept is very popular because of growth of mobile devices, embedded and real time communication, cloud computing and data analytics. The internet of things is a network of physical objects are embedded with electronics, software and sensors having the ability to collect data from the world around us and share data across the internet. The author Geng Wu,Shilpa Talwar,Kerstin Johnsson, Nageen Himyat & Kevin D. Johnson proposed M2m from mobile to internet[6].A connected healthcare environment promotes the quick flow of information and enables easy access to it. Improved home care facilities and regular health updates to clinicians reduce the chances of redundant or inappropriate care, improve patient care and safety, and reduce overall costs of care. Connected health solutions can also be used to track lifestyle diseases such as hypertension, diabetics and asthma which need continuous monitoring. The IoT-MD provides an environment where a patient's vital parameters get transmitted by medical devices via a gateway onto secure cloud-based platforms where it is stored, aggregated and analyzed. It helps store data for millions of patients and perform analysis in real time, ultimately promoting an evidence-based medicine system. On 2011 the author Xiang Zang, Hui-Hong Wang proposed the design and implement of embedded M2M smart home system [7]. M2M communication is something that involves a large number of intelligent machines that share information and make collaborative decisions without direct human intervention. This potentially leads to achieving improved cost efficiency.M2Moffersthetelecommunication industry a great opportunity as it needs a lot of communication systems via various technology families, such as IP, RFID, sensor networks, smart metering, etc. On 2012 the author Takasha Yamanoue, Kentaro oda, koichi shimozono aM2M system using Arduino, Android & wiki software [8] The data generated in a smart grid is more than that generated in a traditional grid due to the continuous two way communication between the parent utility and the smart meter at the

customer's home or business setup. If the infrastructure isn't ready for such communication, it can be a barrier to smart grid deployment. Here, IoT technology plays an important role. It can help streamline the transfer of high-volume data over an internet protocol. The IoT is also needed to establish seamless and effective communication between context aware sensors and the smart meter installed at the user site for automatically switching the devices on or off based on load patterns.On 2022 the author Jeyaselvi ,M., M.Satya, S. Suchitra, S. Jafar Ali Ibrahim,and N. S. Kalyan Chakravarthy [9] proposed the "SVM -Basedcloning and Jamming Attack Detection in IOTsensor Networks". Advances Information Communication Technology and computing. In 2021 Ibrahim, Jafar Ali S., S.Rajasekhar, Varsha,M.Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V.Kumar, and K.J. Kaur [10] proposed "Recent advances in performance and effect of Zr doping with ZnO thin film sensor in ammonia vapour sensing."

PROPOSED METHOD:

This proposed structure can be used in small industry (rubber industry, steel industry, etc.). This post covers the following factors:

a. How to check industrial air quality:
b. To check the number of products manufactured
c. For detecting fires inindustry

i.e.to track the delivery vehicle

The working environment in the industry plays a key role because even if a small percentage of toxic gases are present in the workplace, it can cause health problems for people who oversee the industry and also for people associated with the industry,i.e. the Industrial air quality is monitored with gas detectors built into the Raspberry Pi and the LCD, which actually reduces the number of routine air quality inspections that people in the industry perform.

A lot of work isrequired to verify the quantity ofproducts manufactured in the industry.To reduce labor there are2 IRsensors on the sides of the machines where the products pass between the IR sensors and built into the Raspberry Pi and the LCD screen. Where in the industry can workers be saved for this task.

A fire accident is mainly caused by a malfunction of machines. To detect this fire, use flame sensors built into the Raspberry Pi. Once the fire is detected, the water sprinklers will activate to reduce the crowd Fire.

Real-time latitude and longitude of delivery vehicles can be obtained by integrating GPS (Global Positioning System) with the Raspberry Pi in the delivery vehicle, then saves the received value in the cloud storage and later shows the user the latitude and longitude via an APP to be developed with Android Studio. Use this method in case of a theft of the vehicle, it can be determined using its latitude and longitude.

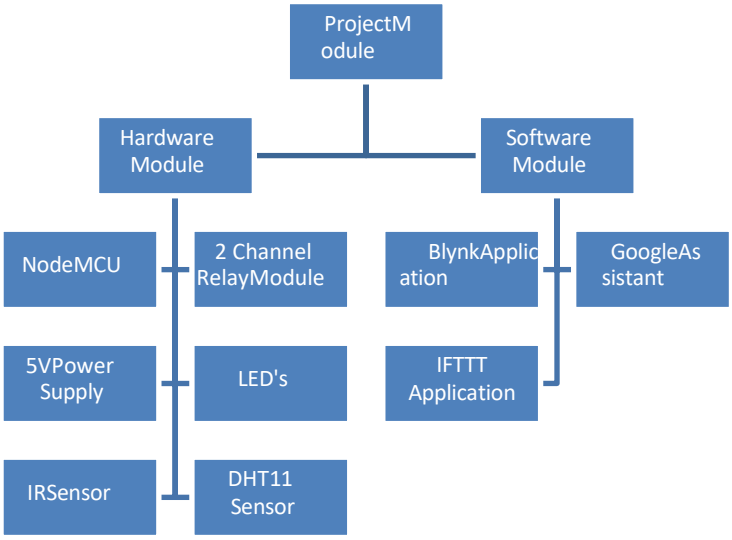


Figure1: Project Layout of Proposed System

NodeMCU: NodeMCU is the microcontroller unit in the prototype. It has a built-in Wi-Fi module (ESP8266), which makes wireless remote switching of home appliances.

Two Channel Relay Module: The Two-channel Relay module consists of 4 individual relays that are physically connected between node MCU and the home appliances. It takes signals from GPIO pins of the node MCU and accordingly connects or disconnects household appliances from the mains. They act as a switching device.

LED: LED are used in this prototype to replace real devices. They show power turndevices on and off. In real-time operation, they would be replaced by actual ones domestic appliances.

Blynk Application: The Blynk application was built for the Internet of Things. It can remotely control hardware, it can display sensor data, save data, visualize, etc. The prototype mainly uses Blynk Application to capture commands from user to hardware over wireless network.

Google Assistant: Google Assistant is a system software that is present on the Android phone. It interprets the voice user commands to turn a device on or off.

IFTTT Application: The voice commands interpreted by the Google Assistant are not understandable therefore, the Blynk application cannot send to the hardware. IFTTT is an intermediate application which interprets commands from the Google Assistant and sends an on and off signal to the Blynk application about Blynk Server.

IR Sensor: An infrared sensor (IR sensor) is a radiation-sensitive optoelectronic component with a spectral sensitivity in the infrared wavelength range 780 nm … 50 μm. IR sensors are now widely used in motion detectors, which are used in building services to switch on lamps or in alarm systems to detect unwelcome guests**.**

DHT 11 SENSOR: The DHT-11 Digital Temperature and Humidity Sensor is a basic, ultra 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).

BLOCK DIAGRAM:-

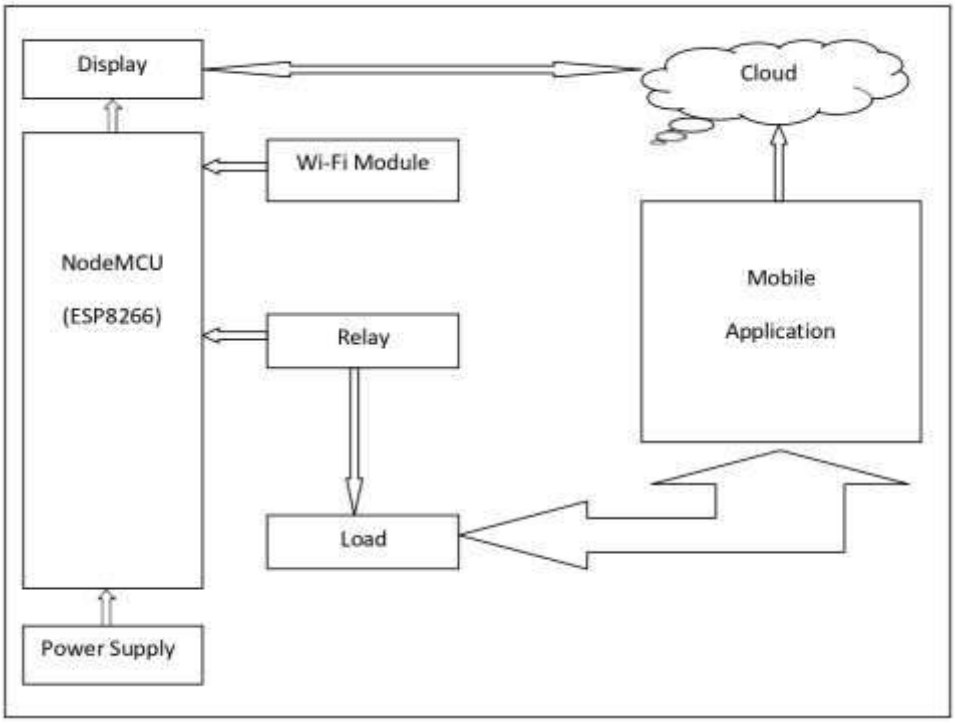


Figure2: Block Diagram of Proposed System

CIRCUIT DIAGRAM:

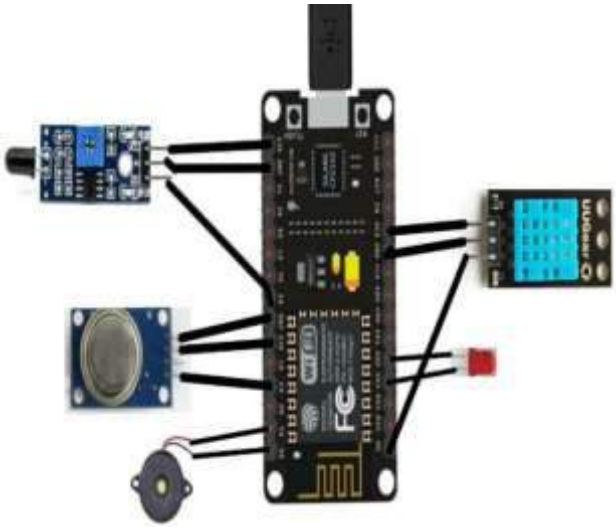


Figure3: Circuit Diagram of Proposed System

WORKING:

The proposed system uses a DHT11 sensor to detect temperature changes and humidity changes. NodeMCU is used to achieve desired operations. The DHT11, fire and smoke sensors are connected to NodeMCU with jumper cables. Whenever the sensors detect unusual changes in their readings, NodeMCU alerts the buzzer. And also sends information to mobile via the Blynk cloud. The whole system is powered by a 9V power supply.

The proposed system has the following functional steps

a. The NODEMCU board is used to dump the program, with the code written in the Arduino IDEprogramming using the injected C encoding.

The core of the circuit is the DHT11 sensor.

b. The SUMME Rare used in this project.

c. The BLYNK cloud is connected to NODEMCU.

d. It detects the temperature and humidity changes in the environment.

Below are the steps to create the proposed system:

The Raspberry Pi is a single-board computer with an ARM processor [approx. 700 MHz], with 256 to 512 MB of RAM memory and with an additional slot for a memory card. Set up and configure the Raspberry Pi by placing it on a non-metallic surface, connect the USB mouse and keyboard to the Raspberry Pi, and later connect the 5V adapter power supply to the Pi and set upEnter the Raspbian operating system.

A flame sensor is activated when it encounters radiation, which the human eye perceives as a yellow-red flame and smoke. Once the Raspberry Pi is configured, connect the flame detectors to the Raspberry Pi's GPIO pins and power the flame sensor is ongiven to the flame sensor of Raspberry Pi. A flame detector is used over a heat detector and a smoke detector because it responds faster and more accurately. When the flame detectors encounter radiation and heat from a point, the flame sensor will become active and send the message to the Raspberry Pi, and later the Raspberry Pi will display the same information on the LCD and activate the water sprinklers to reduce the amount of fire in the industry.

Gas detectors are used to detect flammable gases and toxic gases that leak due to machine malfunction, when these gases cancause health problems

for the people who work and live in the industry. A gas detector is connected to the Raspberry Pi's GPOI pins, and the gasdetector is powered by theRaspberry Pi.When the gas is detected by the gas detector,it will send themessage to the Raspberry Pi and the Raspberry Pi will display the message on the LCD display. With this method, the hassle of regularly checking the air quality can be avoided. An MQ-2 gas sensor can be used.

To check the amount of product going through the machine this can be seen with the 2 IR sensors on the sides of the machine [2] where TSOP1736 can be used driven by a BC557 transistor bias and later send to the Raspberry Pi. The TSOP1736 IR sensor is said to have a photo detection response of 36 KHz.

GPS (Global Positioning System) receivers are used to get real time Latitude and longitude of the vehicle by connecting the GPS to the Raspberry Pi's GPIO pins using TTL (Transistor Transistor Logic) and powering the GPS from the Rasp berry Pi which is3. 3V, the latitude and longitude The delivery vehicle is retrieved and stored in the cloud, and later displayed in the APP developed by Android Studio, where the responsible persons of the delivery team can access the APP and get the real-time value and the estimated time of arrival of the delivery vehicle.

RESULT:

Industrial automation is the control of machinery and processes used in various industries by autonomous systems through theuse of technologies like robotics and computer software.

Industrie simplement automation crease productivity and reduce cost srelated to employees, their benefits and other associated expenses, while increasing precision and flexibility.

With the Industrial Revolution came mechanization, which brought cheaper and more plentiful goods. Generally, the mechanical processes in industries were faster and produced greater quantities of goods but still required skilled workers. Not only did machines require operators but when errors occurred, they would waste materials, cause production issues and even damage equipment.

With the arrival of automation, control loops were added to machine operation. These can be open control loops that allow for human input or closed loops which are fully automated. Industrial control systems (ICS) allow for monitoring and control locally and remotely. With these increasingly advanced control mechanisms, industries can operate 24 hours a day. Productivity has increased, errors are reduced, and quality is improved. However, automation does have some negative impact, including highinitial costs, reduced worker employment and the elimination of some ethical human oversight. As automation continues to advance and gain popularity in new industries, it is possible to see these events increase.

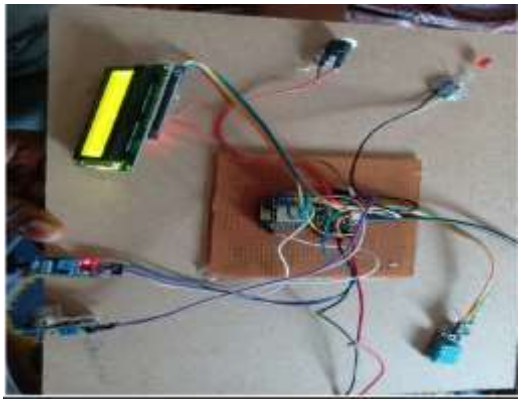


Figure4: Proposed System Implementation

The sensors such as temperature (dht11),humidity (dht11),gass ensors (mq2)are used to sense the parameter values and if they are above the threshold as per the program in raspberry pie it will display on the screen and also updates through blynk app.

Figure5: Displaying the parameters measured on LED screen

In the blynk app the measured parameter are updates such as range of gas is 63ppm and temperature range is32℃ humidity range is 57% for a certain condition. Depends on room temperature the values are changed.

Figure6: Updating of parameters measured in blynk app

CONCLUSION:

This security System is cheaply made from low-cost available components and can be used to control more than others. This system is easily adjustable at any industry or office space. The designed system was tested a number of times and successfully control from different place. Finally, this security system can bealso implemented over Bluetooth, Infrared and WI-FIconnectivitywithoutmuchchangetothedesign. Hence, this system is scalable and flexible.

REFERENCES:

1. Li Da Zu" Internet of Things in Industries: A Survey" IEEE Transactions on Industrial Informatics, vol. 10, no. 4, November 2014
2. Sadeque Reza Khan Professor Dr.M.S.Bhat" GUI Based Industrial Monitoring and Control System``IEEE paper,2014

3. Ayman Sleman and Reinhard Moeller``Integration of Wireless Sensor Network Services into other Home and Industrial networks "IEEE paper.
4. Rajeev Piyareand Seong RoLee``Smart Home-Controland Monitoring System Using Smart Phone"ICCA2013,ASTL Vol. 24, pp. 83 - 86, 2013 © SERSC 2013.
5. Jinsoo Han, Chang-Sic Choi, Wan-Ki Park, Ilwoo Lee Green home energy management system through comparison of energy usage between the same kinds of home appliances 2011 IEEE 15 th International Symposium on Consumer Electronics.
6. Geng Wu, Shilpa Talwar, Kerstin Johnsson,Nageen Himayat, and KevinD. Johnson, Intel(2011):M2M from mobile to Internet.
7. Xiang Yang, Huihong Wang (2011): The design and implement of embedded M2M Smart Home System
8. Takashi Yamanoue, Kentaro Oda, Koichi Shimozono (2012): AM2M system using Arduino,Android and Wiki [Software.
9. Jeyaselvi M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning andJamming AttackDetection in IOT Sensor Networks." Advances in Information communication Technology and computing, pp. 461-471. Springer Singaore, 2022. https://link.springer.com/chapter/10.1007/978-19-0619-0_41.
10. Ibrahim Jafar Ali S., S. Rajasekhar, Varsha, M.Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V.Kumar, and K. J. Kaur. "Recent advancesin performance and effect of Zr doping with ZnO thin film sensor in ammonia vapour sensing."GLOBAL NEST JOURNAL 23, no. 4(2021): 526-531. https://doi.org/10.30955/gnj.004020.

Machine Learning-based Framework for Apple Disease Prediction and Recommendation System: Review & Proposed Architecture

Gulfam Ahamad
Department of ComputerSciences,
Baba Ghulam Shah Badshah University
Rajouri, J&K India
gulfamahamad@bgsbu.ac.in

Owais Ahamad Wagay
Departmentof Computer Sciences,
Baba Ghulam Shah Badshah University
Rajouri, J&K –India
owaisah115@gmail.com

BasitIqbal
Departmentof Computer Sciences,
Baba Ghulam Shah Badshah University
Rajouri, J&K –India
ganiebasit1902@gmail.m

MohdNasee
Department of Computer Sciences, Baba Ghulam Shah Badshah Univerity
Rajouri,J&K India

Abstract: Every year, diseases and pests cost the apple business a lot of money. Farmers face a challenge in identifying distinct apple illnesses since the symptoms of different diseases can be extremely similar and occur at the same time. Here, on the basis of a critical review on apple disease prediction with the help of various computational techniques, we are proposing a Machine Learning (ML) based framework for the effective prediction and prescription of Apple diseases. The architecture of framework uses the Machine learning to

update the knowledge database and automatically or manually recommend the prescription for the identified disease. To smooth the data, the data pre-processing technique will be applied in this framework. Following that, feature extraction will be utilized to choose the features. In the future, the prediction of apple disease and related prescription will be provided to the farmer with the help of various ML techniques such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Convolutional Neural Network (CNN), etc. it would be better to rescue the crop and the farmers' efforts.

Keywords: Machine Learning, Prediction, Prescription, Knowledgebase

I. INTRODUCTION:

The union territory of Jammu and Kashmir is leading in India to export the apple fruit around the globe. Kashmir produces 75% of apples in India. In order to say that around 70% population of Jammu and Kashmir is directly or indirectly dependent on agriculture. In terms of area, apple agriculture covers more than 160000 hectares in Kashmir [1]. The trend to adopt apple horticulture and rapidly switch over to it from agriculture is growing because the apple as a product earns significantly more revenue than other agricultural products despite of more efforts. In terms of numbers, the valley's current apple output is around 12 MTs/hectare. However, this is still significantly lower than the international average of 40 to 60 MT/ha.

One of the most prevalent varieties of fruit growing in India is deciduous fruits. It is a combination of pome (apple & pear) and stone (peach, pulm, apricot, etc.) fruits. The suitable land for the cultivation of aforementioned fruits is mostly Jammu and Kashmir (J&K) and Himachal Pradesh (H.P) in India [2]. The apple is the most significant deciduous fruit in terms of production, consumption, and export. After producing 2.371 million tonnes apples in the year 2017-2018, India got the world's sixth-largest apple producer country [3, 4]. The union territory of Jammu and Kashmir is first to produce the apple in India, followed by H.P, Uttarakhand, and Arunachal Pradesh (Fig. 1 & Fig.2) [3,5]. Almost three million people are employed in the apple industry, either directly or indirectly [6].

Apple is a very productive and as well as widely grown fruit in the world [7]. The quality of the apple fruit attracts the people for the overall development of apple farming as a whole. However, one of the biggest factors affecting the apple quality is the different apple diseases. These diseases have a very negative impact on different aspects of the industry, including the quality and yield of the fruits. In many ways, this directly harms the development of the apple-based economy. Keeping this fact in mind, an important initiative in the form of a precise diagnosis of apple diseases followed up with effective and long-lasting treatments can be underlined as some of the important measures which could be undertaken to reduce or eliminate losses in this sector. In turn, this could promote the overall agricultural sector and hence would help the economy as well.

Unfortunately, the current situation is not encouraging, especially because several aspects of disease care, including as diagnosis and treatment, are left to the farmer. Artificial sickness diagnosis is challenging, especially, when the image characteristics of some infections are same and there is no discernible difference between successive phases of the same infection. This is a technological difficulty. As a result, managing sickness becomes more challenging. Furthermore, due to the unpredictability of illness, many illnesses might go undiscovered for lengthy periods of time, reducing crop quality and yield, and having a direct negative impact on the agricultural economy. As a result, using computational techniques the exact illness detection is becoming increasingly important.

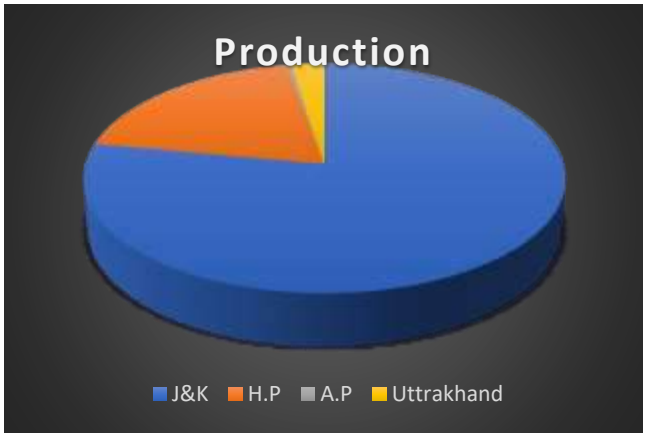


Fig 1: Apple Producing states in India [3,5]

II. LITERATURE REVIEW:

There are many research is going on for the agriculture domain also such as a study is carried out by S Jafar Ali Ibrahim, et. al. [8] on IOT enabled weed controller for typical agriculture. Now, in this article we are going to focus on the apple crops. There are many type of infections observed in apple crops. Some are exists on the leaf and plant of the apple tree and some are seen on or in the apple fruits. In the literature, some studies have been proposed both types of infections in the form of disease. The literature review is classified according to the infection position in apple crop viz. a viz. Leaves, plant and apple fruit.

There are many studies those are reflecting the apple's plant and leaves disease. The various studies are base on identification and classification of apple leaves disease. With the help of deep learning, the identification and classification model for apple's leaves disease is proposed by Asif Iqbal Khan et. al. [9]. They collected a dataset of healthy and sick apple leaves from a variety of orchards around the Kashmir valley for their experiment. Then used it for the proposed model and get some results. The average accuracy rate of model is 97 percent. But there is not any validation and implementation of the model. In same target, Ghai wat et al. [10] experimented to diagnose the illness of plant leaf with various classification techniques. They found that KNN is most appropriate technique for classification purpose and there are some difficulties with SVM due to lack of linear bifurcation in training data. After observing the colour changes of the apple plant leaf, the plant leaf disease is diagnosed. The image processing technique is applied and transformation structure in RGB image is observed [11]. Another study [12] for identifying the apple illness is also based on colour properties of apple. In this study, the reduced feature technique is applied on RGB colour & texture properties of the apples. With this technique the average accuracy of the model is approx. 94 percent. The useful tool for this is BPNN classifier [11].

An image processing technique applied to identify the infected and healthy plants with the help of plant disease pictures. The future work suggested that it will focus on increasing the amount of photographs in the given database and adapting the architecture to the dataset to increase

accuracy. It depicts 17 common illnesses, four bacterial diseases, two mould (oomycete) diseases, two viral infections, and one mite disease [13]. A research work carried out by N. Deb et al [14] on identifying plant disease with image processing. In this work, image-processing applications are utilised to detect crop and plant illnesses early and quickly. Early disease detection allows people who are concerned to treat sick crops or plants, resulting in higher quality and production of the crops and plants cultivated by farmers. This paper discusses recent research, approaches, and contributions provided by many researchers in order to detect plant diseases that impact various plant components.

Gawade, A. (2021) [15], "Early-stage apple leaf disease prediction using deep learning." They presented a system in this work that detects various apple leaf diseases early on and alerts farmers and surrounding research organizations so that suitable control measures can be taken. The data collection contains 1821number of photos of apple leaves, including healthy leaves, diseased leaves, and leaves affected with scab, rust, and other diseases. The suggested regional convolution neural network-based technique is capable of 90 percent accuracy in localising and identifying the condition. A study is carried out by Prakhar Bansal et al.[16] on Apple plant leaves sickness detection with the help of deep CNN technique. In this study, there are 3642 apple leaves in the dataset, divided into four categories: apple scab, apple cedar rust, numerous illnesses, and healthy. On the same dataset, they were trained models. They also looked at the scores of each of the three models separately. Finally, they used averaging to combine the predictions of the three models. On the validation dataset, our final model obtains the 96.25 percent accuracy. This study also shows the future scope as widen the scope of our model's use to cover more types of foliar diseases.

The health of apple plant is also equally important as fruit. A method introduced by Kulkarni Anand H et al. [17] published to recognize the plant ailments in very early stages with the help of Artificial Neural Networks (ANN). It shows the accuracy up to 91 percent. The Gaber filter method is used for feature extraction [18]. The convolutional neural network (CNN) models with deep learning is used to recognise and diagnose plant ailments using simple leaf pictures of healthy and sick plants. After apply many times,

the accuracy of this model was observed approx. 99 percent [17]. By Firasath Nabi and others (2020)[19] employ WSN (Wireless Sensor Network) for precision farming all over the world in this study. It covers a wide range of illnesses that might affect the apple crop. They used a variety of sensors in precision farming applications such as environmental monitoring, precision irrigation, and more. They also conducted a study of plant disease identification techniques for use in a forecasting system.

The apple fruit diseases are very sensitive in respect of last movement reward from the crop. Regarding the apple's fruit disease various studies are carried out. Some of them are given in this article. Shivam Ram Dubey et al. [20] proposed the model for identifying and categorising of the apple fruit diseases. This model uses the K- Means clustering for image segmentation and MC-SVM for classification purposes. The model shows the 93 percent accuracy for classification. A Review had been carried out on Apple Diseases Detection and Classification is carried out by Padaliya Dharm et al. [21]. The various diseases and their influences are addressed. On behalf of that it was proposed the idea to cater the results with the help of mobile application [21]. A technique for identifying the defects in the Golden apple with the help of image processing is introduced. Here, the model mapping the images pixel to pixel with colour contrasts. It is considered healthy tissue if it matches the pixel; otherwise, it is referred to as a defect [22].

A Survey has been done on Apple Fruit Diseases Detection and Classification by Bhavini J. Samajpati et. al. [23] , In this study, it was concluded that the Random Forest Algorithm might be provide the better results with the combining varied colour and texture data sets. According to the normal and abnormal attributes of apple fruits and plants, a discussion carried out in this study. The most highlighted techniques are segmentation, classifications and texture analysis and their pros and cons. In same way a study is carried out by S. Jafar Ali Ibrahim, et. al. on detection and classification of apple fruit disease with the help of image processing techniques [24].

Due to a lack of picture dataset and random occurrence of apple illnesses, the Cycle GAN deep learning technique is used to extract features of

healthy and anthracnose apples, as well as to generate anthracnose lesions on the surface of healthy apples. When compared to traditional picture augmentation methods, this process significantly increases the variety of the training dataset and delivers adequate data for model training [25]. The proposed model shows the better resolution with good prediction. A survey to evaluate the apple's quality is carried out by Jaymit Pandya and Komal Sindhi [26]. In this study, the primary focus is on the examination of grain standard. To begin, the author gathers various qualities such as colour and look, and then groups diseases based on those traits. The database will be created based on disease classification and quality, and disease perception will be based on this information.

While the infection starts in fruit or plant or on leaf, various attributes show the abnormal changes. Focusing on these changes, the fungal infection is identified with the help of SVM and ANN. the k-means technique is applied for the segmentation purpose. The well known fungal infections are leaf blight, leaf spot, powdery mildew, leaf rust, and smut [27]. Scab is also one of the most common apple fruit diseases [30]. The classification accuracy of the ANN classifier ranges from 68.5 percent to 87 percent and average classification accuracies using the SVM classifier were 77.5 to 91.16 percent [28].

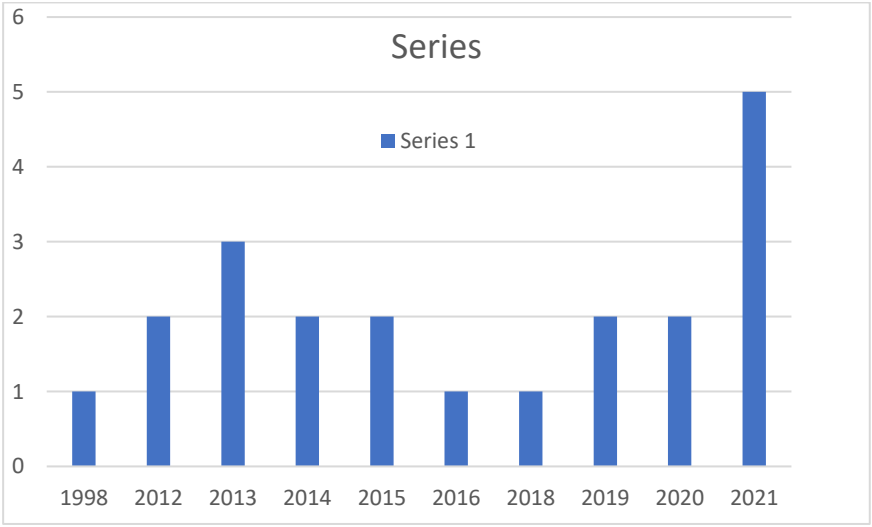


Fig. 2: Statistics for production of Apple in India

III. PROBLEM STATEMENT:

The disease identification in apples is a very difficult in traditional way. There might be information gap in between farmer and pesticide seller. A pesticide seller could not be a perfect doctor for diagnosing the disease. He/she only know about the existing medicines only. There is a need to call the agriculture scientist to visit the orchid and identify the disease and then prescribe the perfect treatment. But it is difficult to call the agriculture scientist for visiting the orchid. There is lack of expertise and unavailability of resources. So, this is a significant problem and farmers are facing and loosing the income.

IV. PROPOSED FRAMEWORK:

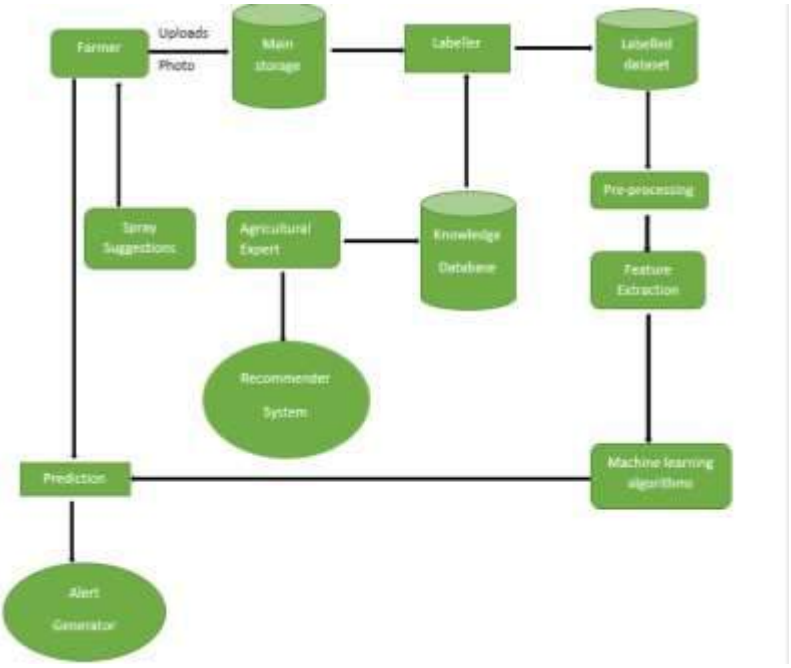


Fig. 3: Proposed Architecture of Framework

1. Data Collection: We will gather the data manually from the orchid of the apple fruits in the form of normal and abnormal apple fruits and plants from a variety of orchids around the Kashmir valley. We will additionally gather data from the internet domain KAGGLE due to a shortage of data and many more sources.

2. Pre-Processing: The image data has various type of impurity due to various technical limitations. So, we have to pre-process the data before using in machine learning models for the accurate results. It's the most crucial step in creating a machine learning model. We don't always come across clean, well-prepared data when working on a machine learning project. Furthermore, each data-related activity must be preceded by the cleaning and preparation of the data.

3. Feature Extraction: The fruit's colour, texture, and form were all taken into consideration. These traits will be used to classify the apple disease. When it comes to correctly recognising apple disease, colour and texture are crucial. Picture segmentation, for example, might be used to detect and diagnose apple scab if specific apples show signs such as a grey or brown corky spot on the fruit. Apple rot, on the other hand, is described as a little sunken, circular brown or black area on the fruit encircled by a crimson halo. The properties of colour, shape, and texture can also be utilised to identify an ailment.

4. Machine: Learing based Model: after pre-processing the data, it will give to the model. Model has a state of art database that is called knowledge base. The entered data will reach in the knowledge base and stored. Now the mapping process will start. If some identified diseases and related prescriptions are available then this ML based model provides the report to the user. Otherwise, the query will put up to the agriculture scientists and then get some identified disease and relevant prescription from the experts' dashboard. Then it will also save in the knowledge base for the same type of query in near future. Now, gradually this system will be reached in mature stage as shown in fig3. Then we can implement it for the farmers. This model uses automation and unsupervised learning approach.

V. CONCLUSION:

The Various research articles are examined in this study with the goal of discovering an effective disease detection method in apples. There are number of reviews and models available in literature those are addressing the

apple disease, types of diseases, infection in the leaves of apple and plants. The theoretical aspects are in plenty but there is lack of practical orientation. There is no model for automatic identification and prescription for apple diseases. To come across this problem, we are proposing the architectural framework for the aforementioned problem. A data set will be prepared in the form of images from the various sources for normal and abnormal condition of the apples. we shall collect the data from a variety of locations around the Kashmir valley. The machine learning models are used to get the better and feasible results. A trained machine is used to detect illness in a new photograph submitted by the farmer, allowing farmers to enhance crop quality and quantity.

REFRENCES:

1. Beigh, M. A, Quadri Javeed Ahmad Peer, Kher , S. K. and Ganai, N. A, "Disease and pest management in apple: Farmers' perception and adoption in J&K state", Journal of Applied and Natural Science 7 (1): 293 – 297 (2015)
2. Bhat KA, Peerzada SH, Anwar A(2015) Alternaria epidemic of apple in Kashmir. Afr J Microbiol Res 9:831-837.
3. HorticulturalStatistics at a Glance (2017) http: /www.agricoop. nic.in/statistics/publication- reports. Accessed May 2019
4. Area and Production of Horticulture Crops-All India(2018) http://nhb.gov.in /statistics/State_Level/2017-18-(Final). Pdf. Accessed May 2019
5. Indian Production of Apple (2018) http://agriexchange. apeda.gov.in /India%20Production/IndiaProductions.aspx?cat=fruiths. Accessed Nov 2018
6. Floods Cause Rs 1000 Cr loss to apple crop in Kashmir: Assocham(2014)http://www.tomesofindia.indiatimes.com.
7. Tian Y, Li E, Liang Z, Tan M and He x (2021) Diagnosis of Typical Apple Diseases: A deep Learning Method Based on Multi-Scale Dense Classification Network.

8. S Jafar Ali Ibrahim, G. Tejaswini, T. Niharika, K. Murari, M. Bhanu Prakash, "IOT Enabled Weed Controller for Typical Agriculture", The International Journal of Analytical and Experimental Modal Analysis, ISSN NO:0886-9367, Volume XIV, Issue III, March/2022: 1-8
9. Asif Iqbal Khan, Saba Banday, S.M.K Qudri, Junaid Latief Shah, Deep Diagnosis: A Real-Time Apple Leaf Disease Detection System Based on Deep Learning in 2021.
10. Jagadeesh D.Pujari, Rajesh Yakkundimath, Abdulmunaf S. Byadgi. Recognition and Classification of Normal and Affected Agriculture Produce using Reduced Color and Texture Features, International Journal of Computer Applications (0975 – 8887) Volume 93 – No.11, May 2014
11. Sanjay B. Dhaygude, Nitin P. Kumbhar. Agricultural Plant leaf disease detection using image processing in 2013.
12. Savita N. Ghaiwat, Parul Arora. Detection and Classification of plant leaf disease using image image processing technique in 2014.
13. V. Suresh, D Gopinath, M Hemavarthini, K Jayanthan, Mohana Krishnan, Plant Disease Detection using Image Processing, International Journal of Engineering Research & Technology (IJERT) in 2020.
14. Epsita Medhi, Nabamita Deb, On Detecting Plant Diseases with Image Processing: A Survey in 2021.
15. Amit Gawade , Subodh Deolekar, and Vaishali Patil, Early-Stage Apple Leaf Disease Prediction Using Deep Learning, Biosc.Biotech.Res.Comm. Special Issue Vol 14 No 5 (2021) Pp-40-43 in 2021.
16. Prakhar Bansal 1 , Rahul Kumar 2 and Somesh Kumar, Disease Detection in Apple Leaves Using Deep Convolutional Neural Network in 2021.
17. Kulkarni Anand H, Ashwin Patil RK. Applying image processing technique to detect plant disease in 2012.
18. Konstantinous P. Ferentinous, Deep learning models for plant disease detection and diagnostics and electronics in agriculture in 2018.

19. Firasath Nabi, Sanjay Jamwal, Kumar Padmanbh, Wireless sensor network in precision farming for forecasting and monitoring of apple disease: a Survey in 2020.
20. S. R. Dubey and A.s. Jalal, " Detection and classification of Apple fruit diseases using complete local binary patterns", in 2012. Third International Conference In Computer and communication Technology.
21. Padaliya Dharm, Pandya Parth, Patel Bhavik, Salkiya Vatsal, Solanki Yuvraj, Mittal Darji,, A Review of Apple Diseases Detection and Classification, International Journal of Engineering & Technology(IJERT), Vol. 8 Issue 04, Apr-2019.
22. V. leemans , H. Magein, and M.F. Destain, "Defect Segmentation in 'Golden Delicious' apples by using color machine vision" , computers and electronics in agriculture, Vol: July 1998.
23. Bhavini J. Samajpati, Sheshang D. Degadwala, A Survey on Apple Fruit Diseases Detection and Classification,International Journal of Computer Applications (0975 – 8887) Volume 130 – No.13, November 2015.
24. S. Jafar Ali Ibrahim, etal. "Detection and Classification of Fruit Disease using Image Processing", Science, Technology and Development, ISSN: 0950-0707, Volume XI Issue II FEBRUARY 2022: 24-31. DOI:21.18001.STD.2022.V11I02.22.352703
25. Yunong Tian, Guogong Yang, Zhe Wang, Enli, and Zi Ze Liang , "Detection of apple Lesions in Orchards based on dep learning methods of CycleGan and YOLOV3-Dense in 08 April 2019.
26. Komal Sindhi, Jaymit Pandya, Sudhir Vegad, Quality evaluation of apple fruit: A Survey, International Journal of Computer Applications (0975 – 8887) Volume 136 – No.1, February 2016.
27. J.D. Pujari,. Color Texture Features are used to classify fungal disease symptoms on cereals. The International Journal of Signal Processing, Image Processing, and Pattern Recognition, vol. 6, no. 3 in 2013.

28. Xiangmimg, “Reaserch on Apple scab (venturia inaequalis) at East Malling Research”, In November 2013.

Collision Avoidance Helping System For Drivers

Mr. M. B. Bheema Kumar[1], M. Kusumanjali[2], G. LathaSree[3], M. Harshitha[4], J. Charanjyothi[5], B.E. Chandrashekhar[6]

[1]Assistant Professor, Department of ECE, GATES Institute of Technology, Gooty.

2,3,4,5,UG student, Department of ECE, GATES Institute of Technology, Gooty.

ABSTRACT:

This paper is used to enhance the driver assistance system for collision avoidance. The advanced driver assistance system (ADAS) is the most innovative technology in the auto mobile industry. Our paper focuses on the problem of road mishaps. Most road accidents occur due to driver negligence. The main purpose of our paper is to propose a method to avoid such accidents using advanced technology at a low cost. We focus on the functions of collision avoidance and liquor detection to give a suitable result for all the problems. We implemented this method using the Arduino UNO, along with the required sensors. The collision avoidance can be reduced by using a cruise control system using an ultrasonic sensor. The system alerts the driver when the oncoming cars are getting too close to the front cars. Using an alcoholic sensor, we can detect the liquor detection of the driver, respectively. With all these functionalities, there is a possibility to avoid accidents. There is future scope for our method by including additional features to reduce accidents.

Keywords: collision, cruise control, Arduino uno

1. INTRODUCTION:

According to the WHO, 1.3 million people die each year due to road accidents. News of road accidents is very common these days. Even though there are some features to overcome accidents using airbags and seat belts. But these can be achieved after road accidents occur. So, we have been able to stop those accidents before they happen. ADAS technology helps drivers with safe driving.

Most road accidents occur due to human mistakes. ADAS schemes are automatic schemes that can be used to decrease the death rates by avoiding human mistakes.

Using this project, we can perform three functions, like collision avoidance. This can be achieved by detecting the object using an ultrasonic sensor. It transmits the sound waves that fall on the object and receives the signal back. If the distance to the object is less than the pre-set value, then the car will automatically slow down. The alcohol consumption of the driver can also be detected using the MQ3 sensor. The LDR sensor is also used to detect the light intensity to determine whether the vehicle exists or not. It can also reduce accidents due to its glaring effect.

2. RELATED WORKS

In 2020, Shivam Kumar et al. proposed a method called the FCW method using CNN. That method can be used to avoid car accidents. It monitors the forgoing cars and measures the distance between the vehicles. If the vehicle is very close, then this method can avoid a collision by giving the audio signal to the driver. The distance can be calculated by the CNN camera, which will be located on the windshield to click rear-side images of the preceding vehicle. After capturing the images of the car, the images will be fed to the model. Next, after feeding those images into the model, it will then detect the distance between the cars. In the past, researchers used technologies like ADAS and FCW Systems, which are entirely dependent on hardware and cost-effective.

In 2020, Ashwini Gadeetal.propose a method called the Intelligent Transport System. They implemented this method using the Raspberry-pi microcontroller through IP connectivity and with the required sensors to control the devices from other locations. They implemented the methods called the auto-lock system and the collision avoidance system. They reduce accidents caused by alcohol consumption, rash driving, and blind spots using these methods. They also protect the driver from unnatural gases present in the vehicle by using required sensors.

In 2020, Gopalakrishnan et al. propose a method called ADAS. They implemented it using RFID technology with the required sensors. Whenever the vehicle enters into an RF region, the speed of the vehicle can be controlled using RFID tags and readers to calculate the distance of front-going cars at a defined speed limit. Required sensors are also used to reduce collisions and crashes with walkers.

In 2019, Ashok Kumar proposed a method called "speed limit recognition using Beacon technology." This technology helps drivers to get notifications when the vehicle entersthe region where beacon tags are placed, and Bluetooth should be enabled on the mobile.

In 2022, S. Jafar Ali Ibrahim, et al., proposed a method called SVM-Based cloning and jamming attack detection on IoT sensor networks. In the proposed technique, the base station can classify the nodes as cloned or normal by checking the distance measurements from IoT devices. The proposed SVM clone shows accurate simulation results with decreasing energy consumption and a false positive rate.

3. PROPOSED SYSTEM:

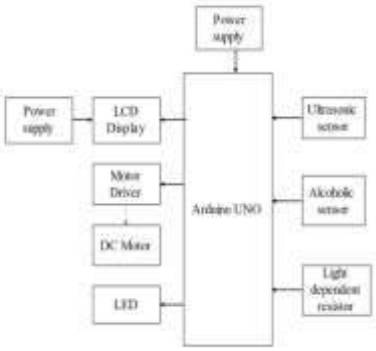


Figure 1. Proposed System Block Diagram

The various components and sensors are used to propose our method, integrated with the Arduino Uno microcontroller. The components are: Arduino Uno, IR sensor, LED, ultrasonic sensor, alcohol sensor, buzzer, LDR (light dependent resistor), LCD display, DC motor, motor driver, and power supply. The details of each sensor can be described as follows:

A. Ultrasonic sensor:

Ultrasonic sensors work by emitting sound waves at a frequency that is too high for human beings to hear. It is a four-pin module with pins for Echo, Trigger, Vcc,and ground, respectively. The sensor will transmit the waves that are reflected off any material, and the waves will be reflected back to the sensor. The reflected wave is then observed by the ultrasonic receiver as shown in fig. The SRF04 sensor provides a 2 cm to 400 cm non-contact measurement function. The ranging accuracy can reach 3mm.

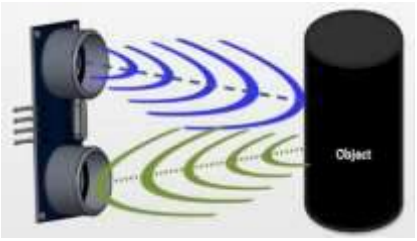


Figure 2: Ultra Sensor Working

Figure3: Ultrasonic Sensor

B. LCD Display:

An LCD display is used to display the outcomes of all sensors. It is better than LED and other displays. It also depletes less power than other displays.

Figure4: LCD Display

C. Alcohol sensor:

An alcohol sensor is used to identify the presence of alcoholic gas in the air and produce an analogue voltage as an output reading. Here we used this sensor to identify whether the driver consumes alcohol or not. If the

driver's consumption of alcohol is over a particular limit, then the system will not be in a power state, so accidents can be reduced. Alcohol sensors are easily interfaced with the Arduino boards and themicrocontroller board.

Figure5: An alcoholic sensor

D. Arduino Uno:

The Arduino Uno is an open-source platform. It is easy to implement both hardware and software, and many Arduino projects can be done easily.

Figure6: Arduino Uno Board

The Arduino board takes the input values from the sensors and actuates the actuators based on the values obtained. It has a CPU along with MU. It sends the output signals to turn on and off the output devices. It can be used in various domains.

E. LDR:

An LDR sensor is used to detect incident light by calculating the intensity of light. The unit gives the output a high when the incident light is present and a low if the incident light is not present. It means by this we can identify whether there is the presence of light or not. If there is light, we can reduce our light intensity there by reducing the accidentsdue to the glaring effect.

Figure7: LDR

3.1 WORKING METHOD:

The micro controller can collect data from the IR sensor, ultrasonic sensor, alcoholic sensor, and LDR sensor. An IR sensor can be used to check whether the vehicle is moving in the same lane or not. It can read the data and check if a vehicle moves to a different lane, then it can track the lane, otherwise it can adjust to the same lane.

The alcohol sensor reads the data and if it is detected that the driver has consumed alcohol, then it immobilises the vehicle. Otherwise, it makes the vehicle free to move.

An ultrasonic sensor reads the data by generating sound waves. It transmits the signal from the transmitter, and the signal falls on the object and reflects the reflected signal back. If the object is detected at a distance of 20m, it again checks if it is present at a distance of less than 3m. If it is, then the vehicle stops moving. Otherwise, the vehicle moves in its normal mode.

If the input value is greater than the threshold value, the light intensity is higher, indicating that a vehicle is present, the low beam is on, and the high beam is off.

4. RESULTS AND DISCUSSION:

Here the figures describe the results. Figure 8 describes the complete implemented module. Figure 9 describes the normal mode of a module whose distance is greater than the predefined threshold. Figure 10 describes the slow mode of the module whose distance is less than the predefined threshold. Figure 11 describes an object nearer to the vehicle with its low beam on. Figure 12 describes the detection of alcohol.

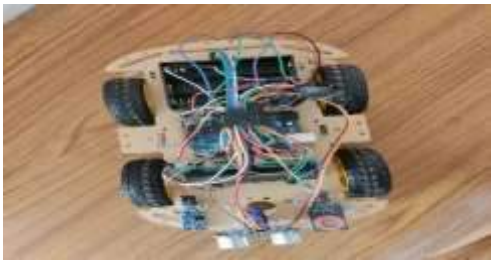


Figure8: Complete Module

Figure 9: Normal Mode

Figure10: Slow Mode

Figure11: Near Vehicle with low beam on

Fig. 12: Detection of Alcohol

5. CONCLUSION AND FUTURE SCOPE:

The adaptive cruise control system is an essential part of our system to control the speed of the vehicle. This system also helps to avoid mishaps from collisions between the vehicles. An alcohol detection system can also be implemented to reduce road accidents. In the future, lane change detection can be included with road surface segmentation. Stereo vision can also be implemented for accurate estimation of front-going vehicles.

REFERENCES:

1. S. Kumar, V. Shaw, J. Maitra and R. Karmakar, "FCW: A Forward Collision Warning System UsingConvolutional Neural Network," 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp.1-5, doi:10.1109/ICE348803.2020.9122858.
2. Ashwini Gade, Rasika Khollam, Karenca Fichardo, "IntelligentTransportation System", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-9, Issue-2, July 2020, pp.590-593,DOI:10.35940/ijrte.B3578.079220
3. K. Gopalakrishnan, S. Afrin Fathima, K. S. Matheshwari, "Advanced Driver Assistance System", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-6, March 2020, pp.3481-3487, DOI:10.35940/ijrte.F8900.038620.
4. W. C. Phoon and P. Y. Lau, "Real-Time Forward Collision Alert System using Raspberry Pi," 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Taipei, Taiwan, 2019, pp. 1-2, doi: 10.1109/ISPACS48206.2019.8986282.
5. H. Elsayed, B. A. Abdullah and G. Aly, "Fuzzy Logic Based Collision Avoidance System for Autonomous Navigation Vehicle," 2018 13th International Conference on Computer Engineering and Systems(ICCES),Cairo, Egypt, 2018, pp. 469-474, doi: 10.1109/ICCES.2018.8639396.
6. S. Sarala, D. S. Yadav, and A. Ansari, "Emotionally Adaptive Driver Voice Alert System for Advanced Driver Assistance System (ADAS)

Applications," in Proceedings of the 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2018, pp. 509–512.

7. S. M. Iranmanesh, H. N. Mahjoub, H. Kazemi, and Y. P. Fallah, "An Adaptive Forward Collision Warning Framework Design Based on river Distraction," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 12, pp. 3925–3934, 2018.
8. Paolo Bosetti, Mauro Da Lio, Andrea, Saroldi, On Curve Negotiation from Driver Support to Automation, IEEE Transaction on Intelligent Transportation Systems, 2015, Volume 16, Issue 4.
9. Chaudhari Priyanka Ramnath, Advanced Driver Assistance System, International Journal of Advanced Research in Electronics and Communication Engineering, 2015, Volume 4, Issue 10.
10. Arun Tigadi, Rudrappa Gujanatti, Anil Gonchi, Advanced Driver Assistance System, International Journal of Engineering Research and General Science, May-June 2016, Volume 4, Issue 3, 2019-2730.
11. B. Karthikeyan, K. Alhaf Malik, D. Bujji Babbu, K. Nithya, S. Jafar Ali Ibrahim, N. S. Kalyan Chakravarthy (2021) "Survey of Cooperative Routing Algorithms in Wireless Sensor Networks", Annals of the Romanian Society for Cell Biology,
12. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IoT Sensor Networks." Advances in Information Communication Technology and Computing, pp. 461-471. Springer, Singapore, 2022.

A Cost-effective Farmer Support System for Better Yield Prediction and Resource Management

Priyanka Prashanth Kumar
Department of Information Science and Engineering
Jyothy Institute of Technology Bangalore, India
priyank aprashanth. kumar@ gmail.com

Sriram Venkatesh
Department of Information Science a Engineering Jyothy Institute of Technology Bangalore, India
sriramvenkatesh28@gmail.com

Dr.Harshvardhan
Tiwari Centre for Incubation, Innovation, Research and Consultancy Jyothy Institute of Technology Bangalore, India tiwari.
Harsh vardhan @ gmail.com

Sriram Venkatesh
Department of Information Science and Engineering
Jyothy Institute of Technology Bangalore, India
sriram venkatesh 28@ gmail.com

Dr.Harshvardhan Tiwari
Centre for Incubation, Innovation, Research and Consultancy Jyothy Institute of Technology Bangalore, India
tiwari. harsh vardhan @gmail.com

Vinay D Bharadwaj
Department of Information Science and Engineering
Jyothy Institute of Technology Bangalore, India
vinay dbharadwaj @ gmail.com

Prajwal G A
Department of Information Science and Engineering Jyothy Institute of Technology Bangalore, India
gaprajwal06@ gmail.com

Abstract : Agriculture is a significant part of India's economy. Agriculture with technology can lead to groundbreaking improvements. Machine learning plays a significant role here, to help predict and formulate essential results. The main focus of this project is to predict and suggest suitable crops to sow in certain soil composition and climatic parameters based on existing historical data collected over years and later subjecting it to Supervised Learning methods and Multi - Class Classification algorithms. So, we propose to create a user-friendly smart phone application that can provide a list of top crops that can be produced in a specific type of soil, and the accuracy of our system can be confirmed using a government website called "Soil Health Card," which collects data on soil composition and its related data.

Keywords: Agriculture, technology, machine learning, soil composition, climatic parameters, supervised learning methods, multi perceptron, multiclass classification algorithms.

INTRODUCTION:

India is the second largest producer of wheat, rice, dry fruits, sugarcane, pulses, and a whole variety of vegetables in the world. Agriculture makes up 20.19% of the Indian GDP as of 2021. But, with limited amounts of land, water, energy and many more resources, it is highly difficult to organize and plan a successful yield for an ever growing population and with the external catastrophes, such as wars, pandemics and what not, the status of the people can stoop down to unimaginable depths.

According to the Department for Promotion of Industry and Internal Trade which has launched a StartupIndia Program [1], it has been observed that Low yield is one of the most pressing concerns confronting India's agricultural sector: India's farm production is 30-50 percent lower than that of wealthy countries.

Low agricultural production is caused by factors such as average farm size, insufficient infrastructure, a lack of utilization of farm technologies and best farming techniques, decreased soil fertility owing to excessive fertilization, and continued pesticide use. Because Indian farms are small, they have limited access to financial services, credit (or lenders), support skills, educational services, and irrigation solutions. In the short term, yield has a direct impact on a farmer's cash flow and ability to adapt to market swings.

In the long-term, yield limits a farmer's ability to invest in their farm's future to increase productivity and reduce risks associated with their crops, but also to invest in their families in areas such as education, healthcare, training, and so on. Farmers are well-versed in their knowledge of agriculture and farming styles, but their knowledge may be limited in terms of only what they were taught from their respective ancestors, preventing them from realizing the full potential of a successful crop yield.

In times like this, technology can be used as a tool in order to bring balance. Machine Learning and Deep Learning methods can be used to build Intelligent Information Systems (IIS). According to a recently published research study [2] that examines the technology characteristics roadmap for a mission-critical intelligent systems world, 80 percent of technological ideas desire intelligent systems that can be considered successful within the next five years.

The proposed solution mentioned throughout the paper inculcates many valuable data resources. The Ministry of Agriculture and Farmers Welfare has provided valuable information about composition of the soil present all over India through their website Soil Health Card [3].

With the help of the internet, there is access to the real time global climatic conditions. Engineers can provide a tool for the public to efficiently organize and plan their resources in order to obtain the maximum yield using an abundance of such data.

In Agriculture, the use of Neural Networks will aid in the prediction of multiple alternative outcomes for the same set of data. The idea is to facilitate the farmers with the right options, which will help them plan better for a good crop yield by effective utilization of resources. The proposed strategy is to

develop a mobile application that can detect a user's location using GPS or manually enter it. The data sources and our dataset include soil and weather data specific to that location. To generate a list of crops that are best suited to grow in a given location, a Machine Learning / Deep Learning model will be built using the best suitable classifiers, such as multi-class classification neural networks. The model will be used to provide the best and next best results in a smart phone application.

LITERATURE SURVEY:

In this section, the work done in previous published papers to predict crop yield and select the best crop is reviewed.

In order to extract and synthesize the techniques and features that have been employed in agricultural yield prediction research, a comparative Systematic Literature Review (SLR) was carried out in this paper [4]. The analysis shows that the most frequently used features in these models are temperature, rainfall, and soil type, and the most frequently used technique is artificial neural networks. The prominent deep learning algorithms used are CNN, LSTM, and DNN. The popular models are Random Forest, Neural Networks, Linear Regression, and Gradient Boosting Tree.

The yield of almost all crops grown in India is predicted by the model created in [5]. The user can anticipate the crop output in the preferred year by using straightforward characteristics like State, District, Season, and Area. In order to estimate the yield, advanced regression techniques including Kernel Ridge, Lasso, and ENet algorithms are utilized, along with a concept known as "Stacking Regression" to improve the algorithms and produce a more accurate prediction.

Linear regression, fuzzy logic, and anfis are all used in this comparative study. The input parameters for the APSim, which was utilized to gather the data, were biomass, esw, radiation, rain, and wheat yield. The RMSE values were used to compare the models' accuracy, and the anfis model outperformed the others [6]. However, as stated in [7], anfis is only appropriate for models with a limited number of input parameters. In this study, artificial neural networks, K-Nearest Neighbors, random forests, support vector machines,

and decision trees are compared. The paper first analyses these models before presenting its own crop selection algorithm for deciding the order of crops to be planted during a season.

This study [8] suggests a creative method of agricultural yield prediction and suggests the ideal environmental variables to maximize crop yield. This forecast will aid farmers in determining the ideal temperature and moisture level for the best crop yield. To predict the crop yield per acre, the method used Multivariate Polynomial Regression, Support Vector Machine Regression, and Random Forest Models. The United States Department of Agriculture's yield and weather data are used in the suggested method. Humidity, yield, temperature, and rainfall are some of the different characteristics in the dataset. At the conclusion, different metrics are calculated to compare the models, including RMSE, MAE, median absolute error, and R-square values.

This research [9] introduces a key framework for the application of machine learning systems in farming. In order to discover which machine learning algorithm is the most accurate at predicting the best harvest for a given field, several methods were compared. The best harvest is the one that outperformed previous years in terms of yield per unit area. The study focuses on six important Bangladeshi crops: potato, wheat, jute, aman rice, aus rice, and boro rice. K-Nearest Neighbor Regression and Multiple Linear Regression (MLR) were employed as algorithms (KNNR). Multiple Linear Regression (MLR), which was used in the analysis, produced the most precise results, and it was added to an Android app.

The suggested system [10] helps farmers choose the best crops based on the sowing season and region. Farmers will gain from it as well because it would boost their net profit. The system develops a model or approach that can suggest a list of crops that is particularly valuable for farmers in their decision-making by taking into account different datasets with respect to five parameters of horticulture data, such as rainfall, temperature, slope, humidity, and soil moisture. This research proposes a useful crop prediction algorithm based on historical data. A pattern matching method is utilized to retrieve the

crop according to region and season after analyzing datasets with Xarray functions.

The majority of the data used in precision agriculture is unstructured because it can be gathered from a variety of sources. Sensors can gather information about farming, including temperature, soil moisture, and nutrient content. While rainfall, heat radiation, and air pressure can all be recorded by weather stations. A statistical yearly report can be used to gather historical data from government colleges, agricultural groups, and websites, such as crop yields from prior years. [11,12,13].

The Modified Recursive Feature Elimination (MRFE), a feature selection technique used by the system [14], facilitates the identification of pertinent parameters from the obtained data set for crop prediction. The MRFE methodology chooses and ranks prominent features using a ranking method. The analysis' findings indicate that bagging is the most precise way for predicting suitable crops, while the MRFE technique is superior for feature selection. The soil properties were taken from Sankarankovil Taluk in Tenkasi, India, while the environmental dataset was obtained from the website of the Tamilnadu Agricultural University. 1000 examples with 16 attributes, 12 of which are soil characteristics, are included in the data set. The other four features are environmental characteristics. The research that was done in this paper used various existing wrapper feature selection techniques such as Boruta, SFFS, RFE, and the proposed MRFE technique.

This article's [15] author uses several machine learning techniques to study crop yield. Combinations are what they are mostly interested in. The testing outcomes demonstrated that, among all the approaches used to implement the model, the XGBoost algorithm is the most demeaning and incomparable. Different kinds of machine learning models can be applied. It is possible to add other crop varieties and geographic areas.

PROPOSED METHODOLOGY:

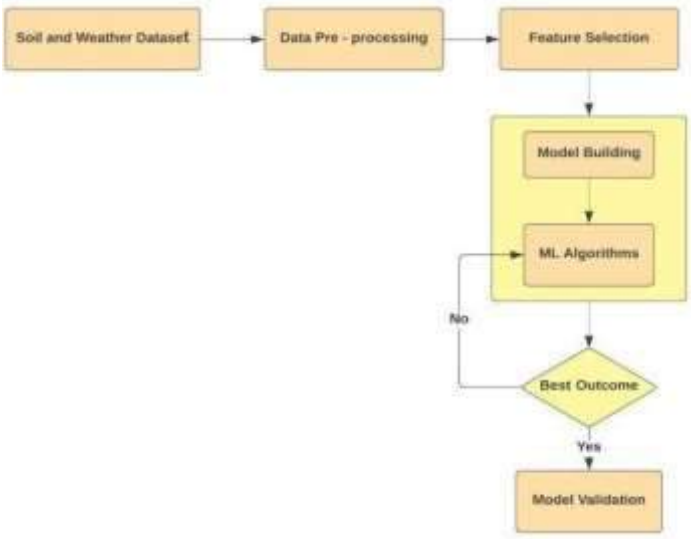


Fig 1. Model Workflow

The proposed system's details are presented in a series of steps (Fig 1). Some of these are:

- Raw dataset collection
- Date Pre – Processing
- Model Building

A. Datasets:

To begin addressing the previously discussed problem, all of the data that represents and determines the successful growth of all of the various crops is necessary. Crop growth is influenced by a variety of factors. The proposed study will consider two major factors: soil composition and climate data.

To take note of soil composition, it is necessary to understand that soil includes parameters such as many minerals, pH levels, moisture levels, and others. Sodium, Potassium, Magnesium, and Nitrogen are a few examples. There are also rainfall levels, humidity, topography, and other climatic factors to consider.

With the above-mentioned elements present in our dataset predicting over 22 different varieties of crops, our system would be able to analyze each

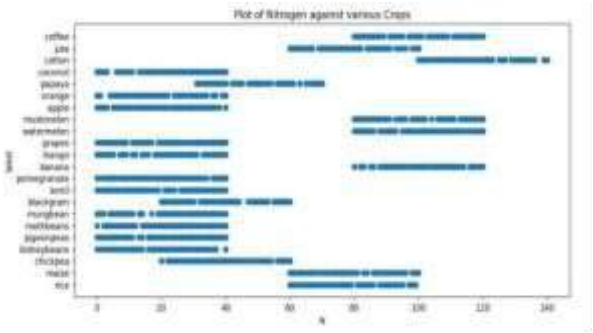
crop across an array of attributes and determine what levels of each element result in a successful yield for that particular crop. Our dataset makes it simpler to determine which crop is best suited for a given set of values for all of the distinct attributes. Fig 2. shows the dataset we have used that contains over 2000 rows worth of data collected and preprocessed later.

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.87974371	82.00274423	6.50298529	202.935536	rice
85	58	41	21.77046169	80.31964408	7.03809636	226.655537	rice
60	55	44	23.00445915	82.3207629	7.84020714	263.964248	rice
74	35	40	26.49109635	80.15836264	6.98040091	242.864034	rice
78	42	42	20.13017482	81.60487287	7.62847289	262.717341	rice
69	37	42	23.05804872	83.37011772	7.0734535	251.055	rice
69	55	38	22.70883798	82.63941394	5.70080568	271.32486	rice
94	53	40	20.27774362	82.89408619	5.71862718	241.974195	rice
89	54	38	24.51588066	83.5352163	6.68534642	230.446236	rice

Fig 2. Structured Dataset used

This set of data may contain various properties required to determine the crops, but the only issue is that the data may be outdated. To have updated values for these characteristics in our dataset, access to sensors and data from across the country would be required. This is a very time-consuming and tedious procedure.

However, there is a government entity; Soil Health Card [3], that provides such soil composition values observed nationwide on their platform. This data can also be accessed in a variety of languages for the convenience of farmers and others. Because the data is updated frequently, our system would be able to formulate predictions almost in real time as it would have access to the most recent data of this level of accuracy. Fig 3. represents the Soil Health Card outlook.



A. Dataset Pre – Processing:

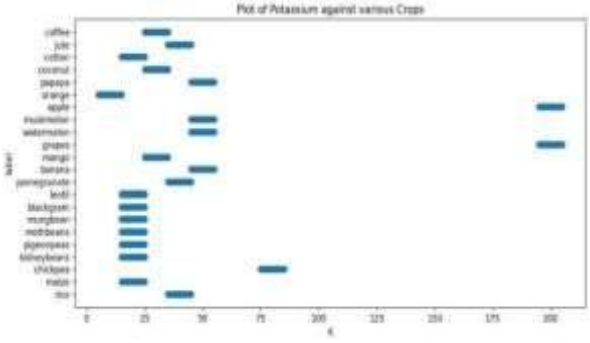
As previously stated, crop yield can be influenced by a variety of factors such as climate, soil, and so on. The data obtained contained a large number of outliers, null values, and discontinuous values. An outlier is a data point that stands out from the rest. Therefore, structuring the data is essential to build our machine learning model. Before data can be used, it must be preprocessed.

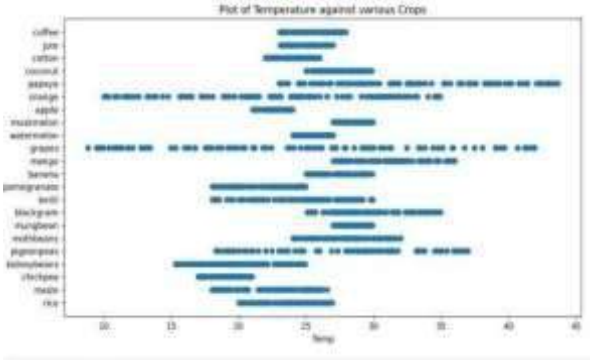
The concept of data preprocessing is the transformation of raw data into a clean data set. Data preprocessing is a vital part in Machine Learning because the quality of data and the useful information that can be extracted

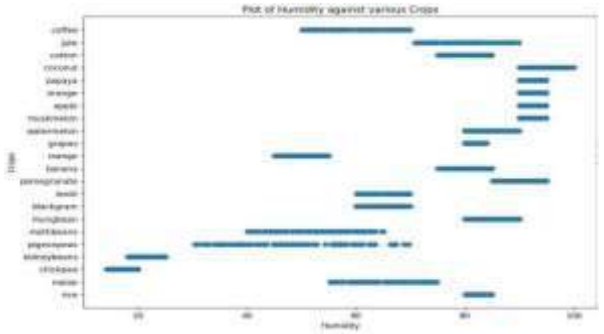
from it directly affects our model's ability to learn; thus, it is critical that we preprocess our data before feeding it into the model. As a result, our model's accuracy and efficiency improve.

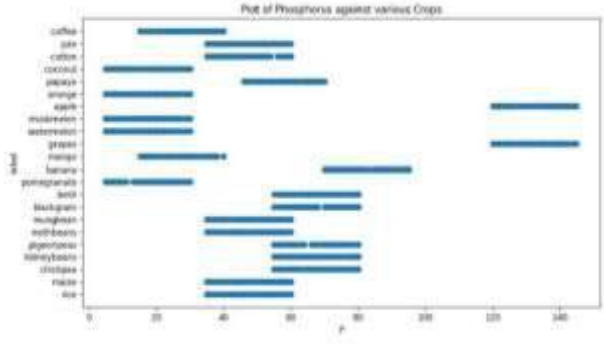
A variety of techniques and methods are applied to the dataset during data preprocessing. For instance, Exploratory Data Analysis(EDA), Data cleaning, Data Integration, Data Transformation etcetera.

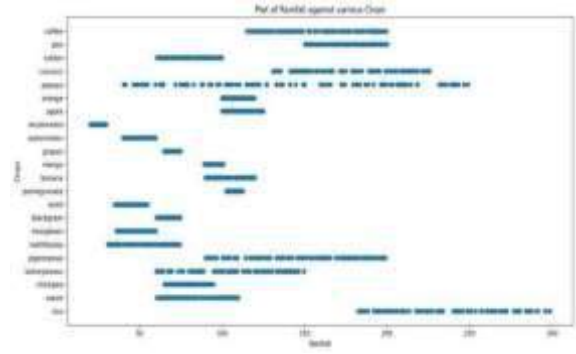
Exploratory Data Analysis: Exploratory Data Analysis is the critical process of performing preliminary investigations on data in order to discover patterns, spot anomalies, test hypotheses, and check assumptions using summary statistics and graphical representations / visualizations before subjecting it to Data cleaning methods. Fig 4. represents the various EDA plots of our dataset.







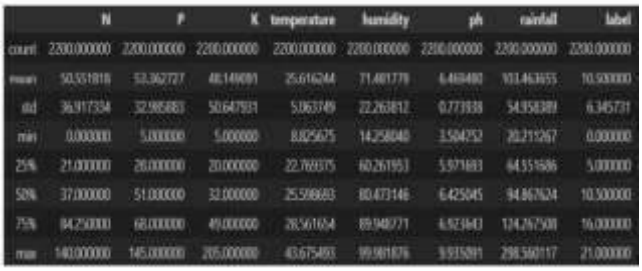




Consider one example from the Temperature plot (Plot 4); the temperature span for grapes to grow can be observed to fall in the range of 0°C to 45°C as per the plot, which covers a wide range of temperatures from cold to hot. This means that grapes are an annual plant, which can be grown all year.

Data Cleaning: Data cleansing is the process of detecting and correcting corrupt or inaccurate records from a dataset. It involves identifying noisy, incorrect, outliers in the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing is also important because it improves data quality and, as a result, overall productivity. When you clean your data, all

outdated or incorrect information is removed, leaving you with only the most suitable data. Fig 5 represents the Descriptive analysis and Outlier determination of our dataset. Dataframe.describe() is used to compute the statistical values for any dataset as seen in Fig 5.



	N	P	K	temperature	humidity	ph	rainfall	label
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481778	6.469480	103.463655	10.500000
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389	6.345731
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267	0.000000
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686	5.000000
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624	10.500000
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508	16.000000
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117	21.000000

Fig 5: Descriptive Analysis

After cleansing the dataset, it is transformed and integrated carefully as per our requirements.

Fig 4. EDA plots of crops against (VS):

- Plot 1: Nitrogen [mg/kg]
- Plot 2: Phosphorus [ppm]
- Plot 3: Potassium [ppm]
- Plot 4: Temperature [C]
- Plot 5: pH
- Plot 6: Humidity [g.kg-1]
- Plot 7: Rainfall [mm]

B. Model Building:

Machine learning models are developed by providing them with functionality that combines many required activities for model development and deployment. Understanding the relationships between machine learning algorithms, machine learning models, and training data is critical. A machine learning model is the result of using training data to train a machine learning algorithm. The accuracy of the model produced by a wide range of machine learning algorithms determines which model is finally used.

Initially, our model was subjected to the KNN, Logistic Regression, XGBoost classifiers in order to predict one specific crop for a given soil composition and weather conditions.

K-Nearest Neighbors: A supervised learning algorithm is the K-nearest neighbors (KNN) algorithm. Its applications include classification and regression problems. It classifies new data points based on feature similarity. The KNN method computes the distance (Euclidean, Manhattan, or Hamming distance) between the test and query data points and the training set data points. After calculating the distance, this method finds the k-nearest neighbors and assigns the new data point to one of the outcome classes using a voting method among neighbors. The model performance has been analyzed with a different number of neighbors, best performance is achieved with 4 neighbors. Euclidean distance is used to measure the distance in our system and the formula to measure it is as follows.

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Logistic Regression. A supervised learning algorithm is the logistic regression method. For binary classification problems, the logistic regression method models the probabilities. It is a more advanced version of the linear regression method. The outcome is based on the use of one or more numerical and categorical predictors. Logistic regression uses the logistic function (Sigmoid function). Logistic regression uses odds instead of proportions. The odds are simply the proportion of the two possible outcomes [16].

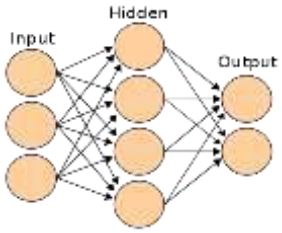
XG Boost. The gradient boosting framework is used in XG Boost, a decision-tree-based ensemble Machine Learning algorithm. Decision tree-based algorithms are now the best for small-to-medium structured/tabular data. It's a well- balanced mix of software and hardware optimization strategies that produce better outcomes with fewer computer resources and in less time.

These models could easily predict a single crop, but the goal is to predict the top crops for a given set of inputs. As a result, multi-class

classification / multi perceptron learning algorithms like ANN come into consideration.

Why is ANN best suited for our proposed model?

Neural networks adapt to changing inputs so that the network generates the best possible result without requiring the output criteria to be redesigned. Fig 6. represents the basic structure of ANN.



Use of One-hot Encoding for multiclass classification using ANN. (Fig 7)

While using the previously mentioned supervised learning algorithms, Label Encoding was sufficient to achieve the highest level of accuracy. For ANN, however, we used a technique known as One hot encoding. For categorical variables with no ordinal relationship, the integer encoding is insufficient. Allowing the model to assume a natural ordering of categories and using this encoding may result in poor performance or unexpected results. In this case, a one-hot encoding can be used to encode the integer representation. The integer encoded variable is removed at this point, and a new binary variable is created for each distinct integer value. It essentially converts categorical label encoded data into binary values, thereby distributing the weightage among the entities equally.

Fig 7: One Hot Encoded data

The most critical step after developing the model is to evaluate it for prediction.The final step of the process of model performance evaluation is

based on its accuracy. The frequency of true negatives, false negatives, true positives, and false positives is represented by the Confusion Matrix. Accuracy, Recall, Precision, F1-Score, G - Mean, Cohen Kappa are some of the important measures to consider while evaluating a model. These are the following formulas to compute the aforementioned parameters.

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{TrueNegative} + \text{FalsePositive} + \text{FalseNegative}}$$

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

$$\text{GMean} = \sqrt{(\text{Recall} * \text{Precision})}$$

The actual labels of test datasets and projected values are used to construct methods which can also be imported from the scikit learn library's metrics module.

RESULTS AND DISCUSSION:

A. Performance Analysis:

To predict one specific crop for a certain soil composition was successful using KNN, Logistic Regression and XGBoost classifiers. The dataset was initially label encoded and results were predicted. Table 1. represents the performance analysis of the various classifiers used. All the statistical parameters are computed as follows.

Table 1. Statistical Performance Analysis and Comparison of classifiers

ML Classifier / *Classification Report*	K- Nearest Neighbors	Logistic Regression	XGBoost
Accuracy	0.9977	0.9975	0.9983
Precision	0.9750	0.97270	0.9820
Recall	0.9750	0.97272	0.98209
F1 - Score	0.9750	0.97270	0.98204
G - Mean	0.9750	0.97270	0.98204
Cohen - Kappa	0.97376	0.96422	0.98121

Table 2. Comparison of Training Score and Model Accuracy

Models	Training Score	Model Accuracy
K- Nearest Neighbors	99.77	97.5
Logistic Regression	99.75	97.27
XGBoost	99.83	99.05

As shown in the above Tables [1 and 2], the XG Boost algorithm was found to be more accurate with about 99.05% accuracy than the other 2 classifiers, but only by a small margin. Fig 8. represents the comparison of accuracies of the classifiers

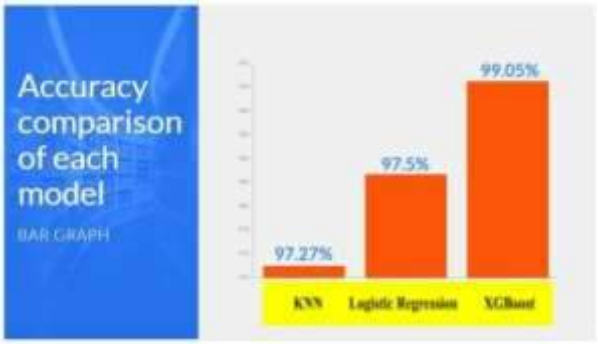


Fig 8: Accuracy Comparison

B. Future Methodologies:

Feature selection techniques are critical for achieving optimal accuracy because they improve the machine learning process and increase the predictive

power of machine learning algorithms by selecting the most important variables and eliminating redundant and irrelevant features. We shall be looking into various feature selection techniques such as the Filter, Wrapper and Embedded methods.

We plan to implement many more machine learning algorithms to determine the accuracy and formulate a comparative study for the same with respect to feature selection techniques. For instance, training the machine learning model with algorithms such as Support Vector Machine, Random Forest, Naive Bayes etc.

To predict multiple crops for the same set of data, we will either use a traditional ANN or build our own custom neural network. This method's main goal is to obtain not only the best but also the next best crops to be sown. Farmers will benefit from this because they will have a wider range of crop varieties to grow without wasting extra farmland, pesticides and many other resources resulting in increased profits and additional benefits.

We plan to incorporate this model into our smartphone application in the future, as well as offer the app in multiple languages, which will be beneficial to farmers who are fluent in their native tongue.

CONCLUSION:

We discussed the importance of crop yield prediction in this paper, as well as how our model could be used to predict multiple crops for the same set of data. As a result, farmers will be able to better manage their resources, resulting in increased profits. We intend to develop and integrate our proposed application into a Smartphone application so that it can be accessed and used at any time. Also, the importance and significance of each step in our model's workflow is discussed. KNN, Logistic Regression and XGBoost algorithms were the supervised machine learning algorithms we used to predict one crop. The XGB classifier produced a higher accuracy of 99.05 percent. We also realized how critical it was to use multi-perceptron methods like ANN in our research. Finally, we can conclude that multi-class classification is critical for predicting multiple outcomes, which is why our system is still in development and will be released soon as a mobile application.

REFERENCES:

1. Department for Promotion of Industry and Internal Trade, Startup India, Govt of India. https://www.startupindia. gov.in/content/sih/en/ agriculture-challenges-det ail.html. (accessed Jan 20, 2022)
2. Characteristics of an Intelligent Systems Future, Forbes (2021). https://www.windriver.com/intelligent-systems. (accessed Jan 20, 2022)
3. Department of Agriculture and Farmers Welfare, Govt of India. https://www.soilhealth.dac.gov.in/. (accessed Jan 20, 2022)
4. Thomas van Klompenburg, Ayalew Kassahun, CagatayCatal. (2020). Crop yield prediction using machine learning: A systematic literature review. In: Computers and Electronics in Agriculture 177 (2020) 105709.
5. Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber (2020). Crop Yield Prediction based on Indian Agriculture using Machine Learning. In: 2020 International Conference for Emerging Technology (INCET) Belgaum, India. IEEE.
6. Aditya Shastry, H. A. Sanjay, Madhura G. Hegde (2015). A parameter based ANFIS model for crop yield prediction. In: 2015 IEEE International Advance Computing Conference (IACC), Bangalore.
7. Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh (2015). Crop selection method to maximize crop yield rate using machine learning technique. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai.
8. Ayush Shah, Akash Dubey, VisheshHemnani, Divye Gala, D. R. Kalbande (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. In: Proceedings of International Conference on Wireless Communication, Lecture Notes on Data Engineering and Communications Technologies 19.
9. Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty (2017). Automated Farming Prediction. In: Intelligent Systems Conference 2017, London, UK.
10. Anjana, Aishwarya Kedlaya K, Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat (2021). An efficient algorithm for predicting crop

using historical data and pattern matching technique. In: Global Transitions Proceedings 2 (2021).

11. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha,
12. V.G.R.R. Devi (2020). Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. In: 5th International Conference on Communication and Electronics Systems (ICCES), 2020.
13. Ephrem Yetbarek, Richa Ojha (2020). Spatio-temporal variability of soil moisture in a cropped agricultural plot within the Ganga Basin, India. Agricultural Water Management 234(140):106108.
14. N. Usha Rani, G Gowthami (2020). Smart Crop Suggester. Advances in Computational and Bio- engineering, Learning and Analytics in Intelligent Systems.
15. Mariammal G, Suruliandi A, Raja SP, Poongothai E (2021). Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers. In: IEEE Transactions On Computational Social Systems.
16. Farnaz BabaieSarijaloo, Michele Porta, Bijan Taslimi, Panos M. Pardalos (2021). Yield performance estimation of corn hybrids using machine learning algorithms, Artificial Intelligence in Agriculture. https://doi.org/10.1016/j.aiia.2021.05.001.
17. Logistiregression,Chapter14.https://nhorton.people.amherst.edu/ips9/IPS09_Ch14.pdf. (accessed Jan 20, 2022)
18. Dr. S. Jafar Ali Ibrahim, Y. Prathyusha, G. Lakshmi, M. Rajini, U. Naveen Sai, "Detection and Classification of Fruit Disease using Image Processing", Science, Technology and Development, ISSN: 0950-0707, Volume XI Issue II FEBRUARY 2022: 24-31. DOI:21.18001. STD.2022.V11I02.22.352703,4-feb2022.pdf (journalstd.com)
19. S Jafar Ali Ibrahim, G. Tejaswini, T. Niharika, K. Murari, M. Bhanu Prakash, "IOT Enabled Weed Controller for Typical Agriculture", The International Journal of Analytical and Experimental ModalAnalytical and Experimental Modal Analysis, ISSN NO:0886-9367, Volume XIV, Issue III, March/2022: 1-8

Arcanum Arcanorum: A secure way of OTP generation

Srishti Lodha
Computer Science and Engineering
Vellore Institute of Technology
Vellore, Tamil Nadu, India
srishti2k1@gmail.com

Abstract: With the increasing importance of digital security, two-factor authentication systems are widely used today. Ideas to make it even more secure would be well received. This paper is mainly concerned with highlighting the problems faced by a lot of the currently used two factor authentication systems, with specific regards to one-time password (OTP) generation and security. After a unique evaluation of some such issues, this paper provides guidelines for developing a secure two-factor authentication mechanism (with a password and mail-delivered OTP system), which generates a truly random, unbiased and non-retrievable unique code. It resolves the discussed possible hacks like duplicate code generation and cracking of probable algorithms, and provides impenetrable security against attacks like dictionary-based and Denial of Service. To break the proposed system and obtain the 6-digit OTP, the attacker will have to brute force 10,00,000 possibilities per second.

Keywords: Authorization, One-time password, Validation, Two-factor authentication, Security issues, Hash.

I. INTRODUCTION:

Arcanum Arcanorum' a Latin phrase, means 'the secret of secrets'. Today, there are hardly any areas where security is not a matter of concern. From banks to military associations, social media, data hiding schemes [32], gaming websites, etc. there is no end to the sectors which need some form of authentication for its users. Initially, the idea of a random string of characters (static passwords) was believed to be a strong security measure to validate user identities. However, with time and advancing technology, cracking passwords started to

become exceedingly simple. While easy ones are guessable and difficult ones are hard to remember, both are vulnerable to attacks like brute force, with varying time to exposure. With standard security systems,it is important to consider the danger of username and password getting known to an outsider/attacker, threatening a client's private information, emails, documents, etc.Consequently, better security measures and stricter authentication systems [24] emerged. The most commonly used technology on these lines is two factor authentication (2FA). However, no matter what system we use for increasing security, its counter attacks are bound to emerge and strengthen with time too. This is why the measures taken should be upgraded consequently until some new technology completely takes over [14, 16].

A. Two-factor authentication systems:

Two step verification or 2FA [10, 11] adds alayer of security, with a secure piece of information added to the single-factor authentication system (username and password),when accessing an application. This added piece is known only to the user, which makes stealing or leaking of personal information, or accessing something as an unauthorized usermore difficultthan it would otherwise be. It is like a token and can be implemented in various forms, including advanced technologies like iris scan and voice/face recognition. The most common token used is a One-Time Password or OTP [13] (refer fig. 1).

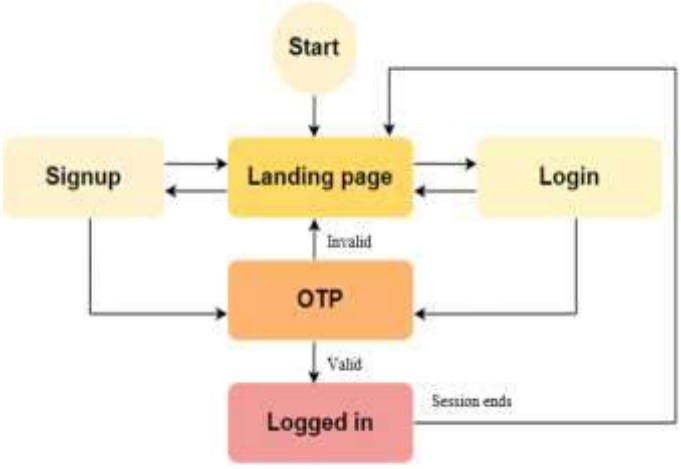


Fig. 1. Working of an OTP based 2FA

B. One-time password:

An OTP isa dynamic password that isvalid for only for one login session on any digital device.We rely on this system to ensure that the person trying to access something (say, an account) is actually who he claims to be. Thus, this OTP is sent to a place which can in principle, be accessed only by the authorized person himself, like the user's personal mail id Consequently, it becomes necessary that this level of security is maintained and no unauthorized person is able to guess or gain access to the OTP in any manner. This provides a more secure way of handling transactions and verifying login procedures.

Conventionally, an OTP is rendered invalid in two cases, when it is successfully used once for validation, or, when it has expired its life span, ideally ranging from a couple of seconds to a few minutes.In this paper, our algorithm is targeted towards a 6-digit OTP, but can easily be extended for a 4-digit one or other types of OTPs by a simple modification (explained later).

The main and required feature of an OTP is that it should be truly random [9, 15] and unique, where each generation should not just be independent of any other generation, but also be completely unpredictable from the result. The algorithm behind the generation should always be a secret. Such a generation is exactly where the challenges arise. In this paper, our improvements are focused on generating such a number. An example of 'truly random' is any case of Bernoulli trials.

C. Layout of the paper:

Vulnerabilities of using OTPs in 2FA systems have been identified and solutions for the same have been proposed in this paper. The study begins with a literature survey (section 2), and problem statement and description (section 3). This is followed by the objective and the methodology (section 4), which serve as a roadmap for the detailed implementation and analysis discussed ahead in section 5. Finally, statistical results in section 6 and conclusion (section 7) wrap up the paper.

II. LITERATURE REVIEW:

Many studies proposed secure OTP generation mechanisms restricted to a particular type of application, like S. A. Lone et al. [18] (solution limited to mobile applications) and H. S. Elganzoury et al. [20] (solution limited to mobile-banking applications). In contrast, the solution in this study is flexible and can be applied with the backend of any application. Several studies like MAR. Karia et al. [12] also highlighted the problems associated with an SMS-based OTP delivery, which has been tackled by us via the use of a mail-based OTP system.

K. S. Hlaing et al. [19] proposed a very straightforward yet interesting RSA-based approach to generate and verify OTP. However, the solution requires that the user has to use the private key with him to decrypt OTP, which is as good as the plain password itself. The entire security depends on the RSA algorithm, so any weaknesses of RSA are valid for this system too. Moreover, DoS and brute-force would still be a threat, which is not the case in this study.

Time-based (TOTP) [27] and HMAC (Hash-based Message Authentication Code)-based (HOTP) [26] systems generate OTP using timestamp and counter parameters. The TOTP system in S. A. El-Booz et al. [21] was not just specific to cloud-platform, but also involved several entities (including third-party auditor and cloud service provider), leading to higher scope for attack and endpoint vulnerabilities. E. P. Nugroho et al. [22] proposed another TOTP system completely based on Advance Encryption Standard algorithm, using current time stamp and a secret phrase. These solutions are prone to forgery and collision attacks. No such attacks have any scope in our study, since no encryption algorithm has been directly implemented and the key is generated independently and uniquely from the database. On similar lines, N. M. Haller [23] suggested an HOTP algorithm (S/KEY authentication system) to specifically fight eavesdropping on network connections. However, as the number of hashing stages in such a system increases (determined by the secret key and encounter), its time and computational complexity also does, and more resources are wasted. The method proposed in Y. Huang et al. [29] again utilizes sequence number and

timestamp to calculate a unique passcode, but the process gives rise to attacks like software cloning, memory scanning and keyboard monitoring.

M. Alzomai et al. [30] proposed mobile phones as scalable OTP devices using trusted computing, which minimized attacks like man-in-the-middle (MITM). Nevertheless, if the attacker is able to masquerade the server, useless OTPs will be produced by the process. The adoptability of the solution is also limited. As per the solution in H. Choi et al. [28], both client and server are supposed to generate the OTP from a challenge value (created by the server), the timestamp and a shared secret key, which is then compared and validated by the server. This approach should reduce MITM attacks, but gives rise to some serious issues like, what if the attacker gets hold of the secret key? If the timestamp by the user and server are not synchronized due to any external or internal issue, even an authenticated user will never be able to pass the security check. Moreover, this entire process adds a burden on the user's end, requiring him to always keep the key handy, and calculating the OTP every time he has to access the application.

In the image-based OTP generation system (with time synchronization) used in N. Vishwakarma et al. [25], all the elements (text and images) used to generate the custom-OTP are user specific and stored in the database itself. This not just makes them easily cached, but the system is highly flawed, because an attacker can easily crack the code if he targets a particular user. Moreover, any compromise on the server side guarantees security breach and the purpose of 2FA is not served. In Ku et al. [6], a Graphical OTP system involving several image-based rounds was proposed, for mobile access of financial services. This system not just requires the users to have more than one device to complete the process, but also burdens them with remembering several alphanumeric codes. None of these problems are faced by the technique proposed in this study and we overcome the possible issues.

III. PROBLEM STATEMENT AND DESCRIPTION:

Methods used for the generation of OTPs may be vulnerable to dictionary, rainbow, and brute force attacks, and cracking of probable algorithms used in the backend. If the server is compromised or the attacker uses mass spam mailing to observe patterns in the OTPs being generated, security is at risk.

A. Duplicate code generation and limited possibilities:

Based on our research of 2FA systems, OTP systems and their apparent disadvantages, weisolated a couple of flaws and developed our algorithms to combat them. One of these is the danger of duplicate code generators. Standard OTP generators usually contain a seed value which when used along with its algorithm, generates an OTP on command, but these can be guessed and the algorithms behind them can be analysed and replicated. Our code combats this issue by generating an OTP using an assortment of hash values and randomly choosing one of them for the later calculations [2], along with implementing TOTP. Let's first understand how such hacks could be carried out in the first place. An attack could be by exploiting the fact that a lot of standard operations always give similar outputs when faced with similar inputs. For example, the modulus operator always gives the same output for a particular number, when the divisor is fixed, like 50 % 9 = 5 (always). Another way to look at this issue is that the output is always going to be from a pool of fixed number of outputs, thus also limiting the range, possibilities and variations of it. For instance, in the above case, no matter the input for the modulus 9 operation, the output is always going to be a number from the set {0,1,2,3,4,5,6,7,8}. As blunt as it sounds, such techniques do form the basis on which some OTP generation algorithms are based, counter-productive to the idea of using secure and independent algorithms.

B. Spam user registration:

This is yet another form of attack most OTP systems are vulnerable to. It involves a hacker logging into the target portal and generating OTPs using a large number of spam accounts. This method also bombards a site's database with spam entries and causes congestion in network traffic, causing Denial of Service (DoS) attacks. The hacker accumulates a large number(more than hundreds and thousands) of OTPs and could possibly guess a pattern in the OTP generation system. This pattern could itself hint the algorithm, or at least, reduce a large number of possibilities from the brute force trials.

Let us take an example to understand this.Hypothetically, say there is a double-digit OTP generation system. If the occurrences of all the generated results are nearly equal, we can predict the possible algorithm is based on the modulus 100 operation. Otherwise, say it is observed from the pool of generated OTPs, that certain numbers are repeated more than the others, while some are not being used at all. The analysis of this result could lead the attacker to some other probable algorithm and eliminate those pseudocodes where the output is always unbiased.Suppose a pool of results contained values only within a certain range.In such a case, it would be reasonable to assume that all the possibilities would be contained in that approximate range, again leading to output prediction.

These are inferences which could be drawn from a large pool of OTP values and is clearly a vulnerability. With the use of our cryptographic suggestions, these kind of attacks are practically non-existent, since it would be extremely difficult to discern a working pattern in an OTP system in accordance to our proposal.

IV. METHODOLOGY:

We propose some suggestions for OTP generation, which overcomes the problems that have been described so far. The proposed system generates a truly random, independent OTP every time a request is made to the server. Our algorithm works hand-in-hand with the database storing the user information. So, a small portion of the entire flow is about adding and fetching information from the database. The next part uses HMAC-SHA1 [3] algorithm, and details with regards to the inputs and output of this part of the entire code. The output from here is a hash which is the input for the next part of the code, along with a separate key. Finally, the last part deals with the generation of the final OTP. Modular arithmetic concepts are used to produce the 6-digit code sent to the user's mail id for validation. Fig. 2 gives an overview of the backend flow, of which the "generate OTP" is this study's focus. This is elaborated in fig. 3.

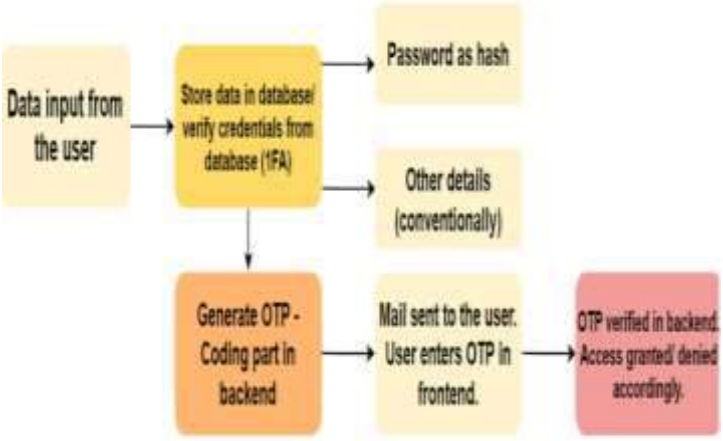


Fig. 2. Overall backend mechanism

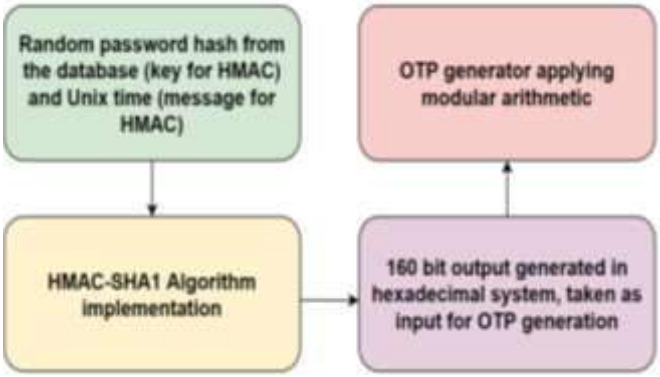


Fig. 3. Flow of control in the backend: 'Generate OTP' section

A. Data set :

Our solution is on OTP generation mechanism that can be implemented in any high-level programming language and be integrated with any application, web-based, mobile-based, etc.

For testing the performance, we needed 100s of user login/signup actions, which of course has to be generated live. We built a basic website and integrated the generation mechanism with its backend, and created about 150 accounts initially (without recording the results). Then, we performed about 700 login actions using the accounts created, and recorded the results obtained (refer section 6).

V. IMPLEMENTATION:

This section gives a detailed explanation of the proposed solution (fig. 3), along with a theoretical analysis.

A. Database system:

Storing all the information that we take from the user in a database as it is, is the conventional approach towards data collection, which is still followed in some real-world applications. Even passwords are collected in the same way. Although user data like name, email etc. can be stored directly (not recommended), passwords should be stored as their hashes. Any hashing algorithm can be applied to convert the password to a hash value [1]. This is an excellent security measure. Hashes are simply irreversible hexadecimal strings with no resemblance to the original string, and of no use to the attacker.

Every time a comparison is made for validation check during login, the user input is converted to its hash and compared with the value stored in the database. If the login password exactly matches the password set during the registration for a particular user, so will their hashes (one of the properties of hashing algorithms). Hence, validation is carried out more or less in the standard approach itself, except comparing the hashes instead of the passwords directly. Hashing is popularly used for storing data today, and is considered as one of the basic security measures. However, there is yet another important reason behind storing passwords as hashes in our algorithm, and this is with regards to the HMAC code, explained in section 5.2.2.

B. HMAC:

Computing the HMAC [8] for a given set of inputs is the first step towards OTP generation according to our proposed algorithm. The output of this step (message authentication code) is calculated by applying hashing over a plain text (the message to be encoded), along with a secret key input. Although any hashing algorithm can be used for this purpose, our demonstration has been done using SHA-1 algorithm, yielding a 160-bit digest of 40 characters (base 16) (see fig. 4).

Fig. 4. Overview of HMAC-SHA1

1. Message input:

By using UNIX time as the message input for HMAC, utterly increased security and randomness in the OTP generation systems is ensured. Now transformed to a time-based process, the Unix time of the system requesting an OTP is converted to a string. This string will change every second (as the system time automatically increments), and so will the message input for HMAC, thus affecting the hash output whatsoever. This process will generate a list of hashes which are always going to look random. An example of a Unix time converted to a string is '16046409' and the same string if generated one second later, would be '16046410', each rendering completely different hashes.

Even if the Unix time is same; say two different systems requesting for an OTP at the exact same moment, chances that the same OTP is sent to both of them are negligible, because another input will affect the output HMAC hash – the secret key.

2. Secret key or k input:

The secret key (equally important for enhancing the security) used for encrypting the message string (here, the Unix Time)involves another continuous random change.

A link is established between the code and the database. However, flow of bulk data does not take place (for obvious security and privacy reasons). Instead, a command shall be run in the backend which retrieves a random password hash from the ones stored in the database, which will be the secret key. Thus, for any generation at the exact same Unix time, the key might differ. The probability of 2 keys being the same decreases exponentially with the increase in the amount of data stored in the database. For example, the probability of getting two same hashes will be much less if the database has a thousand records, compared to if there were only five hundred present. This method has another advantage – it backfires for a hacker trying to implement an attack using a large number of spam generations (section 3.2). More spam

registrations will create more number of hashes stored in the database, thus reducing the chances of exploring a single Unix time input completely in itself.

This method is by far very unique. The combination of an encrypted Unix Time message string and random hash secret key makes our system more secure than other OTP systems. However, there is more to the process. Somehow, if both the key and message are the same despite the above taken measures (which is highly unlikely), the OTP will still differ because of the mathematical calculations we will do on the HMAC output.

3. OTP generation:

As seen above, the HMAC algorithm generates a hash, which is the hexadecimal input used for our OTP generation.

Let us consider a scenario where the HMAC-SHA1 hash is 'B01C36127EDB40C15875F6A7F899F58CAB80B16A'. A variety of mathematical operations can be followed post this step. We suggest one such route.

Procedure:

1) Separate the hexadecimal output into an array with blocks of 2. Hence, the 40-character hexadecimal value yields 20 blocks, stored in 20 indexes of a string array (first block numbered 0, second as 1, and so on till 19).

2) Select the offset as the last character of this string, converted to its decimal equivalent. Select four blocks from the position in this string which equals this offset.

In the considered example, the last character of the string is 'A' in hexadecimal, with a decimal equivalent of 10 - the offset block for this string. Hence, the highlighted characters here form the 4 consecutive blocks which are to be selected: B0 1C 36 12 7E DB 40 C1 58 75 F6 A7 F8 99 F5 8C AB 80 B1 6A

3) Now, select a block from the whole string using a random number between 0 and 19, and another between 0 and 3 to select a block from the 4-block subset we identified.

4) Select the corresponding characters from their respective indexes and concatenate them. For example, if we get random values 16 and 2, then the output string will be ABF8 here; AB contained in the 16th block of the string and F8 contained in the 2nd block from the offset.
5) This string is subsequently reversed and both the strings (original and reversed) are concatenated again:
Original string: ABF8
Reversed string: 8FBA
Resulting string: ABF88FBA (Eq. 1)

Now, we have two strings ready: the first 8-character string we got from the offset (F6A7F899), and the other 8-character string which we just obtained (ABF88FBA). Convert these to their respective decimal values (ie. 4138203289 and 2885193658 respectively).
6) Add these resulting decimal values (=7023396947). Subsequently, take their modulus with 10^6 to obtain a 6-digit unique number. If the number is negative (due to issues concerning the range of data variables), then negate the number. The positive number after negation or directly obtained from the result of the modulus operation is the final OTP. Here, we get 396947. (If we want a 4-digit OTP, the shortest approach here would be a modulus with 10^4).

We used two random numbers to obtain a string which we later add to the blocks extracted from the offset value (steps 3 and 4). The range of the first random number contained 20 values and that of the second had 4. This means that an increase of 20 * 4 = 80 possibilities has been made in this step alone (applying permutations and combinations); 80 options for the string which will be added to the blocks off the offset (step 5), each independent of the other. The randomness and complexity can still be increased. Let us see one such example.

Say, another random number (either 0 or 1) will decide the order of concatenation of the 2 blocks obtained in step 5. From the 4th step, we got two blocks – AB and F8. Instead of concatenating them to form ABF8 straight away, this order can be decided by the number 0 or 1, where 0 could mean ABF8 while 1 would mean F8AB. This improvement will straight away double

the total number of combinations or possibilities; 2*20*4 = 160 = 80*2. Dozens of such possible small improvements will heavily affect the randomness of the final result, making it ultimately almost impossible for the hacker to implement the discussed (and similar) attacking techniques. We want to keep the entire implementation of section 5.3 flexible, while providing a clear picture of our solution.

VI. ANALYSIS AND RESULT:

It has been well explained how the security of password usage has been increased and database independence has been implemented in the previous sections of the paper. The HMAC part of the execution focused a lot on how the OTP generated is always going to be random and how the time-based system is being used to improve the security of 2FA. After analyzing the proposed OTP generating system, the following inferences have been obtained. The figures reflect some important and interesting analysis of the system via graphical representation.

Fig. 5 has been devised after analyzing 100 sample outputs (or OTPs) generated within a span of 5 mins. During the generation, about 150 sample hashed passwords were already present in the database. Hence for these 100 OTPs generated, the Unix time and the input hash for HMAC, and the output HMAC hash, all differed and were random. An analysis with respect to the range within which each OTP lies has been made. The 9 possible ranges are 100000 -199999, 200000- 299999, and so on, up to 999999, superset of which covers all possible 6-digit numbers, one of which will be the OTP.

Fig. 5. Number of OTPs generated within the different ranges

We can see that the difference between the maximum (in the range of 2-3 lacs) and minimum (in the range 6-7 lacs) OTPs generated (16 and 7 respectively), is only of 9 OTPs, which is very less. Hence, the mean deviation is very less and all the values lie near the average of the number of OTPs per range. Hence, the generated OTPs are truly random and do not favor any particular range/values, neither is any range excluded from the pool of possibilities. Additionally, the aim of this graph is to prove that the algorithm produces unbiased, random result during each generation. OTPs have nothing to do with any other generation and are totally independent. From the inferences drawn, we can see that there is not a huge variation between the OTPs generated in any 2 ranges, thus fulfilling the aim.

In fig. 6, 600 OTPs are examined. The message input taken for HMAC has been externally fixed for all these generations, to examine the probability of generating the same OTPs at the exact same instant. Table 1 helps in understanding the result. Number of occurrences was 1 for 507 OTPs, which is equivalent to 507/600 *100% ~ 84% of the total codes generated. Similarly, 32 OTPs occurred twice, therefore (32*2)/600 *100 ~ 11%, and so on. Hence, with just 150 entries in the database during the testing, 84% of the generated OTPs were unique even with the same message input for HMAC! Thus, we can safely conclude that with a realistic and high number of entries stored in the database, it is highly unlikely that two users will receive the same OTs at the same instant.

Table I. Performance evaluation for fixed Unix time

Number of occurrences of an OTP	Number of such OTPs	Number of unique OTPs
1	507	507
2	64	32
3	21	7
4	8	2

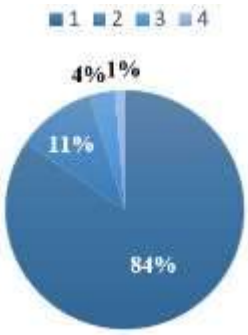


Fig. 6. OTP occurrences for fixed Unix time

VII. CONCLUSION:

Software tokens are not just favored as more secure, but also widely used, making them a probable area for hackers to explore. Out of many such web-attacks [17], this paper targets a few important ones, like mailers, spam, backdoors, DoS and Distributed DoS, database hack, brute force etc. Despite the fact that 5G and next-generation cryptography [4, 5] is being extensively researched upon, OTPs are a reality and much in demand when it comes to current security systems, especially with respect to online accounts. Therefore, it is never too late to improve the existing mechanisms, even if the target is just a small flaw with the OTP generating systems. A few of such flaws, like duplicate code and spam user generation, and cyber-attacks stated above have been the quarry of this paper and appropriate solutions have been proposed with respect to 2FA.

Concepts of hashing, time stamping, HMAC and modular arithmetic were used in the proposed system. From the statistical analysis, we saw that the OTPs generated are truly random and do not favor any particular pool of vales. The chances of an OTP to repeat (when generated at the same instant) exponentially reduces with more entries in the database. It is as low as 17% with just 150 registrations. If time is also varied (even by 1 second), the chances of two OTPs being the same is reduced to less than 1%. In addition, all the problems discussed in sections 2 and 3 are also tackled. To break the system and obtain the OTP, the attacker will have to brute force 10,00,000 possibilities per second. Therefore, we can conclude that the chances for the same are negligible. The scope of this study can be extended by using Internet

time in place of Unix time. The extraction of Internet time is comparatively challenging, but enhances the security and purpose of the paper. Finally, more levels of security using advanced and next-gen cryptography techniques can be introduced to this system.

REFERENCES:

1. Pittalia PP (2019) A comparative study of hash algorithms in cryptography. Int J Comp Sci Mob Comput 8(6): 147-152.
2. Kale MA, Dhamdhere S (2018) Survey Paper on Different Type of Hashing Algorithm. Int J Adv Scient Res Eng Trends 3(2).
3. Sahu A, Ghosh DSM (2017) Review Paper on Secure Hash Algorithm with its Variants. Int J Tech Innovation Mod Eng Sci 3(5). http://dx.doi.org/10.13140/RG.2.2.13855.05289
4. Gunathilake N, Buchanan WJ, Asif R (2019) Next Generation Lightweight Cryptography for Smart IoT Devices: Implementation, Challenges and Applications. Inst Electrical and Electronics Engineers 707-710. https://doi.org/10.1109/WF-IoT.2019.8767250
5. Kaul V, Nemade B, Bharadi DV, Narayankhedkar DSK (2019) Next Generation Encryption Using Security Enhancement Algorithms for End to End Data Transmission in 3G/4G Networks. ScienceDirect 79: 1051 – 1059. https://doi.org/10.1016/j.procs.2016.03.133
6. Ku Y, Choi O, Kim K et al (2013) Two-factor authentication system based on extended OTP mechanism. Int J Comp Math 90(12). http://dx.doi.org/10.1080/00207160.2012.748901
7. Amin A, Haq IU, Nazir M (2017) Two Factor Authentication. Int J Comp Sci and Mob Comput 6(7).
8. Boppudi K (2011) Efficient HMAC Based Message Authentication System for Mobile Environment. Glob J Comp Sci Tech 11(19).
9. Kang SY, Lee IY (2007) A Study on Secure R (Random digit)- OTP Scheme suing Random Digit. Proc Korea Inf Processing Soc Conf 14(2): 1254-1257.
10. Cristofaro ED, Du H, Freudiger J, Norcie G (2014) A comparative usability study of Two-Factor authentication. arXiv:1309.5344v2

11. Chen AQ, Goh W (2015) Two Factor Authentication Made Easy. Int Conf Web Eng 9114: 449-458. http://dx.doi.org/10.1007/978-3-319-19890-3_29
12. Karia MAR, Patankar DAB, Tawde MP (2014) SMS-based one time password vulnerabilities and safeguarding OTP over network. Int J Eng Res Tech 3(5): 1339-1343.
13. Malikovich KM, Turakulovich KZ, Tileubayevna AJ (2019) A Method of Efficient OTP Generation Using Pseudorandom Number Generators. Int Tech 1-4. https://doi.org/10.1109/ICISCT47635.2019.9011825
14. Wang D, He D, Wang P, Chu CH (2014) Anonymous Two-Factor Authentication in Distributed Systems: Certain Goals Are Beyond Attainment. IEEE TransDependable 12(4). https://doi.org/10.1109/TDSC.2014.2355850
15. Lee YL, Lim H , Lee H (2010) A Study on Efficient OTP Generation using Stream Cipher with Random Digit. Adv Comm Tech (ICACT) 2.
16. Abdullahi Y, Muazu MB, Adedokun AE (2017) Response Time Improvement on One Time Password (OTP) Technique to Prevent Replay Attack in a Radius Environment. British J App Sci Tech 19(1).
17. Kaur D, Kaur P (2016) Empirical Analysis of Web Attacks. Procedia Comp Sci 78: 298-306. https://doi.org/10.1016/j.procs.2016.02.057
18. Lone SA, Mir AH (2021) A novel OTP based tripartite authentication scheme.IntJPervasive Comput Comm.https://doi.org/10.1108/IJPCC-04-2021-0097
19. Hlaing KS, Aung NA (2019) Secure One Time Password (OTP) Generation for user Authentication in Cloud Environment. Int J Trend Scient Res Dev.
20. Elganzoury HS, Abdelhafez AA, Hegazy AA (2018) A new secure one-time password algorithm for mobile applications. 2018 35th National Radio Sci Conf (NRSC) 249-257. https://doi.org/10.1109/NRSC.2018.8354370
21. El-Booz SA, Attiya G, El-Fishawy N (2015) A secure cloud storage system combining Time-based One Time Password and Automatic Blocker

Protocol. 2015 11th Int Computer Engineering Conference 188-194. https://doi.org/10.1109/ICENCO.2015.7416346

22. Nugroho EP, Putra RRJ, Ramadhan IM (2016) SMS authentication code generated by Advance Encryption Standard (AES) 256 bits modification algorithm and One time Password (OTP) to activate new applicant account. 2016 2nd Int Conf Sci Inf Tech (ICSITech) 175-180. https://doi.org/10.1109/ICSITech.2016.7852629
23. Haller NM (1994) The s/key (tm) one-time password system. Symposium on Net Dist Sys Sec 151-157.
24. Al Qahtani AAS, El-Awadi Z, Min M. (2021) A Survey on User Authentication Factors. 2021 IEEE 12th Annual Inf Tech, Electronics Mob (IEMCON) 0323-0328. https:// doi.org/10.1109/IEMCON 53756.2021.9623159
25. Vishwakarma N, Gangrade K (2016) Secure Image Based One Time Password. Computer Science Department, Rajiv Gandhi Technical University, Shri Ram Institute of Technology Jabalpur, India.
26. Raihi DM, Bellare M, Hoornaert F et al (2005) Hotp: An hmac-based one-time password algorithm. The Internet Society, Net Working Group. RFC4226.
27. Raihi DM, Machani S, Pei M, Rydell J (2011) Totp: Time-based one-time password algorithm. Internet Req for Comments, 685E.
28. Choi H, Kwon H, Hur J (2015) A secure OTP algorithm using a smartphone application. 2015 Seventh Int Conference on Ubiquitous and Future Networks 476-481. IEEE. https://doi.org/10.1109/ICUFN.2015.7182589
29. Huang Y, Huang Z, Zhao H, Lai X (2013) A new one-time password method. IERI Procedia, 4: 32-37. https:// doi.org/10.1016/ j.ieri.2013.11.006
30. [30] Alzomai M, Jøsang A (2010) The mobile phone as a multi OTP device using trusted computing. 2010 Fourth Int Conf Net System Sec 75-82. IEEE. https://doi.org/10.1109/NSS.2010.39
31. Jafar Ali Ibrahim. S, Mohamed Affir. A "Effective Scheduling of Jobs Using Reallocation of Resources Along With Best Fit Strategy and

Priority", International Journal of Science Engineering and Advanced Technology(IJSEAT) – ISSN No: 2321- 6905, Vol.2, Issue.2, Feb-2014, http://www.ijseat.com/index.php/ijseat/article/view/62

32. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352-357 http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf

Deep Learning Approach for UAV-Based Wildfire Detection and Segmentation

K. Sreenath[1], G. Hemalatha Reddy[2], K. Venkata Ranga Reddy[3], G.SnehaSandhya[4], MD. Mahaboob Basha[5]

[1]Assistant Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Abstract: Wildfires are a worldwide natural disaster causing important economic damages and loss of lives .Early detection and prediction of fire spread can help reduce affected areas and improve firefighting.Unmanned Aerial Vehicles were employed to tackle this problem due to their HIGH flexibility, their low-cost, and their ability to cover wide areas during the day or night.However, they are still limited by challenging problems such as small fire size, background complexity, and image degradation.To deal with the limitations, we adapted and optimized Deep Learning methods to detect wildfire at an early stage. A novel deep ensemble learning method, which combines EfficientNet-B5 and DenseNet-201 models, is proposed to identify and classify wildfire using aerial images.

Keywords: Efficient Net-B5 and Dense Net-201 models.

1. INTRODUCTION:

Forest fire accidents are one of the most dangerous risks due to their frightening loss statistics. The fires cause human, financial, and environmental losses, including the death of animals and the destruction of wood, houses, and million acres of land worldwide. In 2021, forest fires have occurred in several countries such as the European Union countries, the US (United States), central and southern Africa, the Arabian Gulf, and South and North America . They affect 350 million to 450 million hectares every year. In the western US alone, the frequency of wildfires and the total area burned increased by 400% and 600%, respectively, in the last decade. In addition, approximately 8000

wildfires affected 2.5 million hectares each year in Canada. Generally, wildfires are detected using various sensors such as gas, smoke, temperature, and flame detectors. Nevertheless, these detectors have a variety of limitations such as delayed response and small coverage areas. Fortunately, the advancement of computer vision techniques has made it possible to detect fire using visual features collected with cameras. However, traditional fire detection tools are being replaced by vision-based models that have many advantages such as accuracy, large coverage areas, small probability of errors, and most importantly the ability to work with existing camera surveillance systems. Through the years, researchers have proposed many innovative techniques based on computer vision in order to build accurate fire detection systems. In recent years, Unmanned Aerial Vehicles (UAV) or drone systems were deployed in various tasks such as traffic monitoring, precision agriculture, disaster monitoring , smart cities , cover mapping , and object detection . They are also very practical and well developed for wildfire fighting and detection. UAV-based systems help with precise fire management and provide real-time information to limit damage from fires thanks to their low cost and ability to cover large areas whether during the day or night for a long duration. The integration of UAVs with visual and/or infrared sensors help in finding potential fires at daytime and night time. Furthermore, fire detection and segmentation showed impressive progress thanks to the use of deep learning (DL) techniques. DL-based fire detection methods are used to detect the color of wildfire and its geometrical features such as angle, shape, height, and width. Their results are used as inputs to the fire propagation models. Thanks to the promising performances of DL approaches in wildfire classification and segmentation , researchers are increasingly investigating this family of methods. The existing methods use input images captured by traditional visual sensors to localize wildfire and to detect the precise shape of fire; they achieved high results. However, it is not yet clear that these methods will perform well in detecting and segmenting forest fire using UAV images, especially in the presence of various challenges such as small object size, background complexity, and image degradation. To address these problems, we present in this paper a novel deep ensemble learning method to detect and classify

wildfire using aerial images. This method employs Efficient Net-B5 and Dense Net-201 models as a backbone for extracting forest fire features. In addition, we employed a deep model (Efficients eg) and two vision transformers (Trans U Net and Trans Fire) in segmenting wildfire pixels and detecting the precise shape of fire on aerial images. Then, the proposed wildfire classification method was compared to deep convolutional models (MobileNetV3-Large - Small , DenseNet169 , EfficientNet-B1-5 , X ception [28,29], and InceptionV3), which showed excellent results for object classification. Trans U Net, Trans Fire, and Efficients eg are also evaluated with U- Net More specifically, three main contributions were reported in this paper. First, a novel DL method was proposed to detect and classify wildfire using aerial images in order to improve detection and segmentation of wildland fires. Second, vision transformers were adopted for UAV wildfire segmentation to segment fire pixels and ident fire.

1.2. Related Works:

DL approaches are employed for fire detection and segmentation using aerial images. They proved their ability to detect and segment wildfires . They can be grouped into three categories: DL approaches for UAV-based fire classification, DL approaches for UAV-based fire detection, and DL approaches for UAV-based fire segmentation.

1.2.1. Fire Classification Using Deep Learning Approaches for UAV Images

Convolutional Neural Networks (CNNs) are the most popular AI models for images classification tasks. They extract feature maps from input images and then predict their correct classes (two classes in our case: Fire and Non-Fire). Three main types of layers, which are convolutional layers, pooling layers, and fully connected layers, are employed to build a classical CNN architecture:

- Convolution layers are a set of filters designed to extract basic and complex features such as edges, corners, texture, colors, shapes, and objects from the input images. Then, activation functions are used to add the non-linearity transformation. It helps CNN to learn complex features in the input data.

Various activation functions were employed, such as Rectified Linear Unit (ReLU) function , Leaky ReLU (LReLU) function [31], parametric ReLU (PReLU) function , etc.

- Pooling layers reduce the size of each feature map resulting from the convolutional layers. The most used pooling methods are average pooling and max pooling.
- The fully connected layer is fed by the final flattened pooling or convolutional layers' output, and the class scores for the objects present in the input image are computed. CNNs showed good results for object classification and recognition . Motivated by their great success, researchers presented numerous CNN-based contributions for fire detection and classification using aerial images in the literature, and these are summarized.

Table 1. Deep learning methods for UAV-based fire classification.

Ref.	Methodology	Smoke/Flame	Dataset	Accuracy (%)
[34]	CNN-17	Flame/Smoke	Private dataset: 2100 images	86.01
[35]	AlexNet	Flame	Private dataset: 23,053 images	94.80
	GoogLeNet			99.00
	Modified GoogLeNet			96.90
	VGG13			86.20
	Modified VGG13			96.20
[24]	Xception	Flame	FLAME dataset: 48,010 images	76.23
[36]	Fire_Net	Flame	UAV_Fire dataset: 1540 images	98.00
	AlexNet			97.10
[29]	VGG16	Flame	FLAME dataset: 8617 images	80.76
	VGG19			85.43
	ResNet50			88.01
	InceptionV3			87.21
	Xception			81.30
[37]	Fog computing and simple CNN	Flame	Private dataset: 2964 images	95.07
[38]	Fire_Net	Flame/Smoke	Private dataset: 2906 images	97.50
	AlexNet			95.00
	MobileNetv2			99.30

Chen et al. proposed two CNNs to detect wildfire in aerial images. The first CNN contains nine layers. It consists of a convolutional layer with Sigmoid function, max-pooling layer, ReLU activations, Fully connected layer, and Softmax classifier. Using 950 images collected with a six-rotor drone (DJI

S900) equipped with a SONYA7 camera, the experimental results showed improvements in accuracy compared to other detection methods . The second includes two CNNs for detecting fire and smoke in aerial images [34]. Each CNN contains 17 layers. The first CNN classifies Fire and Non-Fire, and the second detects the presence of smoke in the input images. Using 2100 aerial images, great performance (accuracy of 86%) was achieved, outperforming the first method and the classical method, which combines LBP (Local Binary Patterns) and SVM . Lee et al. employed five deep CNNs, which included AlexNet ,GoogLeNet , VGG13 , a modified GoogLeNet, and a modified VGG13 to detect forest fires in aerial images:

- AlexNet includes eleven layers: five convolutional layers with ReLU activation function, three max-pooling layers, and three fully connected layers;
- VGG13 is a CNN with 13 convolutional layers
- GoogLeNet contains 22 inception layers, which employ, simultaneously and in parallel, multiple convolutions with various filters and pooling layers;
- Modified VGG13 is a VGG13 model with a number of channels of each convolutional layer and fully connected layers equal to half of that of the original VGG13;
- Modified Google Net is a Google Net model with a number of channels of each convolutional layer and fully connected layer equal to Google Net and the modified Google Net achieved high accuracies thanks to data augmentation techniques (cropping, vertical, and horizontal flip). They showed their ability in detecting wildfires in aerial images Shamsoshoara et al. proposed a novel method based on the exception model for wildfire classification. Exception architecture is an extension of the Inceptionv3 model with the modified depth-wise separable convolution, which contains 1 × 1 convolution followed by a n × n convolution and no intermediate Re LU activations. Using 48,010 images of the FLAME dataset and data augmentation techniques (horizontal flip and rotation), this method achieved an accuracy of 76.23%. Treneska et al. also adopted four deep CNNs, namely InceptionV3, VGG16, VGG19, and Res Net50, to detect wildfire in aerial images. ResNet50 achieved the best accuracy with 88.01%. It outperformed

InceptionV3, VGG16, and VGG19 and the recent state-of-the-art model, exception, using transfer learning techniques and the FLAME dataset as learning data. Srinivas etal. also proposed a novel method, which integrates CNN and Fog computing to detect forest fire using aerial images at an early stage. The proposed CNN consists of six convolutional layers followed by the Re LU activation function and max-pooling layers, three fully connected layers, and a sigmoid classifier that determines the output as Fire or Non-Fire. This method showed a great performance (accuracy of 95.07% and faster response time) and proved its efficiency to detect forest fires. Zhao etal . presented a novel model called "Fire Net" to extract fire features and classified them as Fire and Non-Fire. Fire Net is a deep CNN with 15 layers. It consists of eight convolutional layers with ReLU activation functions, four max-pooling layers, three fully connected layers, and a softmax classifier. Using the UAV_Firedataset ,Fire_Net achieved an accuracy of 98% and outperformed previous methods. Wu etal. used a pretrained MobileNetv2 model to detect both smoke and fire. MobileNetv2 is an extended version of Mobile Ne tv 1 , which is a lightweight CNN with depth-wise separable convolutions. It requires small data and reduces the number of parameters of the model and its computational complexity. It employs inverted residuals and linear bottlenecks to improve the performance of MobileNetv1. Using transfer learning and data augmentation strategies, this method achieved an accuracy of 99.3%. It outperformed published state-of-the-art methods such as Fire Net and Alex Net and proved its suitability in detecting forest fire on aerial monitoring systems.

1.2.2. Fire Detection Using Deep Learning Approaches for UAV :

Region-based CNNs are used to detect, identify, and localize objects in an image. They determine the detected objects' locations in the input image using bounding boxes. These techniques are divided into two categories: one-stage detectors and two-stage detectors. One-stage detectors detect and localize objects as a simple regression task in an input image. Two-stage detectors generate the ROI (Region of Interest) in the first step using the region proposal network. Then, the generated region is classified and its

bounding box is determined. Region-based CNNs showed excellent accuracy for object detection problems. They are also employed to reveal the best performance in detecting fires on aerial images. Table 2 presents deep learning methods for UAV-based fire detection. Jiao etal. exploited the one-stage detector, YOLOv3 , to detect forest fires. YOLOv3 is the third version of YOLO deep object detectors. It was proposed to improve the detection performance of older versions by making detections at three different scales and using the Darknet-53 model, which contains 53 convolutional layers as a back bone . Testing results revealed great performances and high speed Jiao et al. also proposed the UAV-FFD (UAV forest fire detection) platform, which employs YOLOv3 to detect smoke and fire by using UAV-acquired images. YOLOv3 showed high performance with reduced computational time (F1-score of 81% at a frame rate of 30 frames per second). It proved its potential in detecting smoke/fire with high precision in real-time UAV applications . Alexandrov et al.adopted two one-stage detectors (SSD and YOLO v2) and a two-stage detector (Faster R-CNN [56]) to detect wildfires. Using large data of real and simulated images, YOLOv2 showed the best performance compared to Faster R-CNN, SSD, and hand-crafted classical methods. It proved its reliability in detecting smoke at an early stage . Tang et alalso presented a novel application to detect wildfire using 4K images, which have a high resolution of 3840 × 2160 pixels collected by UAS (Unmanned Aerial Systems). A coarse-to-fine strategy was proposed to detect fires that are sparse, small, and irregularly shaped. At first, an ARSB (Adaptive sub-Region Select Block) was employed to select subregions, which contain the objects of interest in 4K input images. Next, these subregions were zoomed to maintain the area bounding box's size. Then, YOLOv3 was used to detect the objects. Finally, the bounding boxes in the subregions were combined. Using 1400 4K aerial images, this method obtained a mean average precision (mAP) of 67% at an average speed of 7.44 framesper second.

Table 2. Fire detection using Deep learning methods for UAVs

Ref.	Methodology	Smoke/Flame	Dataset	Results (%)
[56]	YOLOv3	Flame	Private dataset: 2,860,000 images	F1-score = 81.3
[53]	YOLOv2	Smoke	Private dataset: 12,000 images	Accuracy = 98.3
	Faster R-CNN			Accuracy = 95.9
	SSD			Accuracy = 91.1
[57]	YOLOv3	Flame/Smoke	Private dataset: 2,860,000 images	F1-score = 81.3
[57]	YOLOv3 and ABSS method	Flame	Private dataset: 1,860 K images	mAP = 67.3

1.2.3. Fire Segmentation Using Deep Learning Approaches for UAV:

Image segmentation is very important in computer vision. It determines the exact shape of the objects in the images. With the progress of deep learning models, numerous problems were tackled and a variety of solutions was proposed with good results. Deep learning models are also used to segment fire pixels and detect the precise shape of smoke and/or flame using aerial images. Table 3 shows deep learning methods for UAV-based fire segmentation. For example, Barmpoutis et al. proposed a 360-degree remote sensing system to segment both fire and smoke using RGB 360-degree images, which were collected from UAV. Two DeepLab V3+ models that are encoder–decoder detectors with ASPP (Atrous Spatial Pyramid Pooling) were applied to identify smoke and flame regions. Then, an adaptive post-validation scheme was employed to reject smoke/flame false-positive regions, especially regions with similar characteristics with smoke and flame. Using 150 360-degree images of urban and forest areas, experiments achieved an F1-score of 94.6% and outperformed recent state-of-the-art methods such as DeepLabV3+. These results showed the robustness of the proposed method in segmenting smoke/fire and reducing the false-positive rate [58]. Similarly to wild fire classification, Shamsoshoara et al. proposed a method based on the encoder–decoder U-Net for wildfire segmentation. Using a dropout strategy and the FLAME dataset, U-Net obtained an F1-score of 87.75% and proved its ability to segment wildfire and identify the precise shapes of flames .Frizzi et al. also proposed a method based on VGG16 to segment both smoke and fire.

This method showed good results (accuracy of 93.4% and segmentation time per image of 21.1 s) using data augmentation techniques such as rotation, flip, changing brightness/contrast, crop, and adding noises. It outperformed previous published models and proved its efficiency in detecting and classifying fire/smoke pixels. Sensors 2022, 22, 1977 6 of 1

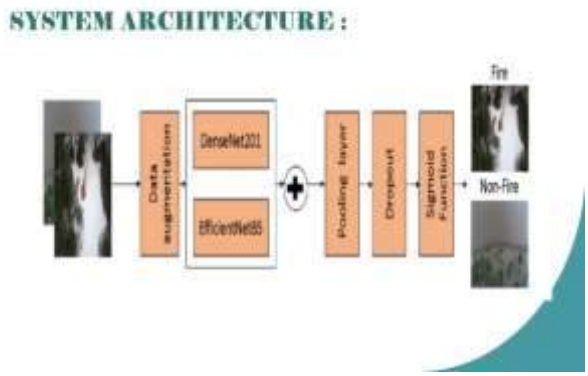
Table 3. Fire segmentation using deep learning methods for UAVs.

Ref.	Methodology	Smoke/Flame	Dataset	Results (%)
[58]	DeepLabV3+	Flame/Smoke	Fire detection 360-degree dataset:	F1-score = 81.4
	DeepLabV3+		150 360-degree images	F1-score = 94.6
	+ validation approach			
[60]	U-Net	Flame	FLAME dataset: 5137 images	F1-score = 87.7
[61]	U-Net	Flame/Smoke	Private dataset: 366 images	Accuracy = 90.2
	CNN based on VGG16			Accuracy = 93.4

2. PROPOSED SYSTEM:

To deal with the limitations, we adapted and optimized Deep Learning methods to detect wildfire at an early stage. A Deep Learning ensemble method, which combines EfficientNeB5 and DenseNet-201 models, is proposed to identify and classify wildfire using aerial imageThe obtained results are promising and show the efficiency of using Deep Learning and vision transformers for wildfire classification and segmentation. It proved its ability in classifying wildfire even small fire areas.

Data Collection - FLAME dataset (Fire Luminosity Airborne-based Machine learning Evaluation)
Data Processing (Handling missing data, filtering, Train-Test Split)
Model Selection (Deep Learning - EfficientNet-B5 and DenseNet-201)
Train and Evaluate the Model.
Prediction and Classification (Wild fire)



DenseNet201: DenseNet-201 is a convolutional neural network that is 201 layers deep. We can load a pretrained version of the network trained on more than a million images from the ImageNet database.

Efficient Net is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolutioncoefficient.

3. Dataset:

In the area of deep learning, many large datasets are available for researchers to train their models and perform benchmarking by making comparisons with other methods. However, until recently, there was a lack of a UAV dataset for fire detection and segmentation. In this work, we use a public database called FLAME dataset (Fire Luminosity Airborne-based Machine learning Evaluation) [45] to train and evaluate our proposed methods. The FLAME dataset contains aerial images and raw heat-map footage captured by visible spectrum and thermal cameras onboard a drone. It consists of four types of videos, which are a normal spectrum, white-hot, fusion, and green-hot palettes. In this paper, we focus on RGB aerial images. We used 48,010 RGB images, which are split into 30,155 Fire images and 17,855 Non-Fire images for wildfire classification task. Figure 5 presents some samples of the FLAME dataset for fire classification. On the other hand, we used 2003 RGB images and their corresponding masks for fire segmentation task. Figure 6 illustrates some examples of RGB aerial images and their corresponding binary masks

Table 4. Dataset subsets for classification.

Dataset	Fire Images	Non-Fire Images
Training set	20,015	11,500
Validation set	5003	2875
Testing set	5137	3480

Figure 5. Examples from the FLAME dataset. Top line: Fire images and bottom line: Non-Fire images.

4. Conclusions:

In this paper, A novel ensemble learning method, which combines EfficientNet-B5 and DenseNet-201 models, was developed to detect and classify wildfires. Using the FLAME dataset, experimental results showed that our proposed method was the most reliable in wildfire classification tasks, presenting a higher performance than recent state-of-the-art models.

References

1. Aytekin, E. Wildfires Ravaging Forestlands in Many Parts of Globe. 2021. Available online: https://www.aa.com.tr/en/world/ wildfires-ravaging-forestlands-in-many-parts-of-globe/2322512 (accessed on 20 November 2021).
2. Dimitropoulos, S. Fighting fire with science. Nature 2019, 576, 328–329.
3. Westerling, A.L.; Hidalgo, H.G.; Cayan, D.R.; Swetnam, T.W. Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. Science 2006, 313, 940–943.
4. Canadian Wildland Fire Information System. Canada Wildfire Facts. 2021. Available online: https://www.getprepared.gc.ca/ cnt/hzd/wldfrs-en.aspx (accessed on 20 November 2021).

5. Gaur, A.; Singh, A.; Kumar, A.; Kulkarni, K.S.; Lala, S.; Kapoor, K.; Srivastava, V.; Kumar, A.; Mukhopadhyay, S.C. Fire Sensing Technologies: A Review. IEEE Sens. J. 2019, 19, 3191–3202. [CrossRef]
6. Ghali, R.; Jmal, M.; SouideneMseddi, W.; Attia, R. Recent Advances in Fire Detection and Monitoring Systems: A Review. In Proceedings of the 18th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Genoa, Italy, 18–20 December 2018; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; Volume 1, pp. 332–340.
7. Gaur, A.; Singh, A.; Kumar, A.; Kumar, A.; Kapoor, K. Video flame and smoke based fire detection algorithms: A literature review. Fire Technol. 2020, 56, 1943–1980.
8. Dao, M.; Kwan, C.; Ayhan, B.; Tran, T.D. Burn scar detection using cloudy MODIS images via low-rank and sparsity-based models. In Proceedings of the IEEE Global Conference on Signal and Information Processing GlobalSIP), Washington, DC, USA, 7–9 December 2016; pp. 177– 181.
9. Töreyin, B.U.; Dedeo ˘glu, Y.; Güdükbay, U.; Çetin, A.E. Computer vision based method for real-time fire and flame detection. Pattern Recognit. Lett. 2006, 27, 49–58.
10. Zhang, J.S.; Cao, J.; Mao, B. Application of deep learning and unmanned aerial vehicle technology in traffic flow monitoring. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 9–12 July 2017
11. S. Jafar Ali Ibrahim et al, "An Overview on Network Representation", Learning Journal of University of Shanghai for Science and Technology ISSN: 1007-6735 Vol.23, Issue 01 Page 60-69, January 2021, https://jusst.org/wp-content/uploads/2021/01/Network-Representation-Learning.pdf

Detecting Chronic Kinley Disease By Using A Voting Classifier

M.Rama [1], G.Tejaswini [2], T.Niharika [3], K.Murari [4], M.BhanuPrakash[5]

Assistant Professor (IT), Dept. of Information Technology, QIS College of Engineering & Technology, India(1).

Final Year (B.Tech), Dept. of Information Technology, QIS College of Engineering & Technology, India (2, 3, 4, 5)

Abstract:

Chronic Kidney Disease (CKD) is a global health problem with high morbidity and mortality rate, and it induces other diseases. Since there are no obvious symptoms during the early stage of CKD, patients often fail to notice the disease. Early detection of CKD enables patients to receive timely treatment to ameliorate the progression of this disease. Machine learning models can effectively aid clinicians achieve this goal due to their fast and accurate recognition performance. In this, we proposed a machine learning methodology for diagnosing CKD. The CKD data set was obtained from the University of California Irvine (UCI) machinelearning repository, which has a large number of missing values. In general we have six machine learning algorithms like logistic regression, random forest, support vector machine, k-nearest neighbour, naive Bayes, classifier and feed forward neural network to establish models. In existing system random forest achieved the performance with 99.75% diagnosis accuracy. In proposed system, By analysing the misjudgments generated by the above model, we proposed an integrated model that combines logistic regression and random forest by using perceptron which could achieve an efficient accuracy of 99.83% after ten times of simulation.

Hence, we speculatedthat this methodology could be applicable to more complicated clinical data for disease diagnosis.

Introduction:

1. Background:

Chronic Kidney Disease is a global health problem affecting approximately 10% of the world population. The percentage of prevalence of CKD in China is 10.8% and the range of prevalence is 10%-15% in the United States. According to another study, this percentage has reached 14.7% in the Mexican adult general population. This disease is characterised by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious symptoms in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has highmorbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. CKD is a progressiveandirreversible pathologic syndrome. Hence, the prediction and diagnosis of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease.

Machine learning refers to a computer program, which calculates and deduces the information related to the task and obtains the characteristics of the corresponding pattern. This technology can achieve accurate and economical diagnoses of diseases; hence, it might be a promising method for diagnosing CKD. It has become a new kind of medical tool with the development of information technology and has a broad application prospect because of the rapid development of electronic health records. In the medical field, machine learning has been used to detect human body status, analyze the relevant factors of the disease and diagnose various diseases. For example, the models built by machine learning algorithms were used to diagnose heart disease, diabetes and retinopathy, acute kidney injury, cancer and other diseases. In these models, algorithms based on regression, tree, probability, decision surface and neural network were often effective. In the field of CKD diagnosis, Hodneland et al. utilized image registration to detect renal morphologic changes. Vasquez-Morales et al. established a classifier based on neural networks using large-scale CKD data, and the accuracy of themodel on their test data was 95%. In addition, most of the previous studies utilized the CKD data set that was obtained from the UCI machine learning repository. Chen et al. used k-

nearest neighbour (KNN), support vector machine (SVM) and soft independent modelling of class analogy to diagnose CKD, KNN and SVM achieved the highest accuracy of 99.7%. In addition, they used fuzzy rule-building expert system, fuzzy optimal associative memory and partial least squares discriminant analysis to diagnose CKD, and the range of accuracy in those models was 95.5%- 99.6%. Their studies have achieved good results in the diagnosis of CKD. In the above models, the mean imputation is used to fill in the missing values and it depends on the diagnostic categories of the samples. As a result, their method could not be used when the diagnostic results of the samples are unknown. In reality, patients might miss some measurements for various reasons before diagnosing. In addition, for missing values in categorical variables, data obtained using mean imputation might have a large deviation from the actual values. For example, for variables with only two categories, we set the categories to 0 and 1, but the mean of the variables might be between 0 and 1. Polat et al. developed an SVM based on feature selection technology, the proposed models reduced the computational cost through feature selection, and the range of accuracy in those models was from 97.75%-98.5%. J. Aljaaf et al. used novel multiple imputation to fill in the missing values, and then MLP neural network (MLP) achieved an accuracy of 98.1%. Sub a set al. used MLP, SVM, KNN, C4.5 decision tree and random forest (RF) to diagnose CKD, and the RF achieved an accuracy of 100%. In the models established by Boukenze et al., MLP achieved the highest accuracy of 99.75% . The studies focus mainly on the establishment of models and achieve an ideal result. However, a completeprocess of filling in the missing values is not described in detail, and no feature selection technology is used to select predictors as well. Almansour et al. used SVM and neural network to diagnose CKD, and the accuracy of the models was 97.75% and 99.75%, respectively. In the models established by Gunarathne etal., zero was used to fill out the missing values and decision forest achieved the best performance with the accuracy was 99.1%.

To summarize the previous CKD diagnostic models, we find that most of them suffering from either the method used to impute missing values has a limited application range or relatively low accuracy. Therefore, in this work, we propose a methodology to extend application range of the CKD diagnostic models. At the same time, the accuracy of the model is further improved.

2. Problem Statement:

Chronic kidney disease (CKD) is a global public health problem affecting approximately 10% of the world's population [1], [2]. The percentage of prevalence of CKD in China is 10.8% [3], and the rangeof prevalence is 10%-15% in the United States [4]. According to another study, this percentage has reached 14.7% in the Mexican adult general population [5]. This disease is characterised by a slow deterioration in renal function, which eventually causes a complete loss of renal function. Nowadays, health care industries are providing several benefits like fraud detection in health insurance, availability of medical facilities to patients at inexpensive prices, identification of smarter treatment methodologies, construction of effective healthcare policies, effective hospital resource management, better customer relation, improved patient care and hospital infection control.

Disease detection is also one of the significant areas of research in medicine.

3. Objectives

- To check Weather the person is a effected with Chronic Kidney Disease or not
- To check the stage of the affected person
- Comparing the model with another model

However, for the bias, when the updating method of (14) was used, the obtained decision line could classify the samples, but the line was located at the edge of the solution area, so it is not reliable. To solve this problem, a new bias adjustment strategy proposed in chapter 4 of the previous literature was referred to. The proposed models reduced the computational cost through feature selection, and the range of accuracy in those models was from 97.75%-98.5%. J. Aljaaf et al. used novel multiple imputation to fill in the missing values, and then MLP neural network (MLP) achieved an accuracy of 98.1%. Subas et al. used MLP, SVM, KNN, C4.5 decision tree and random forest (RF) to diagnose CKD, and the RF achieved an accuracy of 100%. In the models established by Boukenzeet al., MLP achieved the highest accuracy of 99.75%.

The studies focus mainly on the establishment of models and achieve an ideal result. However, a complete process of filling in the missing values is not described in detail, and no feature selection technology is used to select predictors as well. Almansour al. used SVM and neural network to diagnose CKD, and the accuracy of the models was 97.75% and 99.75%, respectively. In the models established by Gunarathne et al., zero was used to fill out the missing values and decision for est achieved the best performance with the accuracy was 99.1%.

Related Work:

Classification techniques, Feature selection,and Ensemble model are the most significant and vital tasks in machine learning and data mining. A lot of research has been conducted to apply data mining and machine learning classification technique, feature selection method and ensemble model on different medical datasets to classify disease datasets. Many of them show good classification accuracy.

Polat, H e al. [6] Diagnosis chronic kidney disease using SVM and effective feature selection methods. They used wrapper and filter feature selection method to reduce the dimensionality of the feature. In their work they improve accuracy by implanting SVM without feature selection the accuracy rate was 97.75%, SVM with the classifier subset evaluator combine with greedy stepwise the accuracy rate was 98%, SVM with the wrapper subset evaluator combine with a best first search engine the accuracy rate was 98.25, SVM with the classifier subset evaluator combine with greedy stepwise the accuracy rate 98.25. And finally, SVM with the filter subset evaluator combine with best first search the accuracy was 98.5.

Bashir, S. et al [15] they proposed ensemble classifier which uses majority Vote Based framework for prediction of heart disease. They used five heterogeneous classifiers used to construct the ensemble model. The classifiers are Naïve Bayes, decision tree based on Gini Index, decision tree based on information Gain, memory based learner and SVM. After experiment using stratified cross-validation show that their MV5 framework has achieved an accuracy 88.5% with 86.96% sensitivity, 90.83% specificity and 88.85 F-

Measure and they compare with the base classifiers show to increase the average accuracy of the ensemble model. They involved proposing the ensemble approach. The first approach generates the individual classifier decision and the second approach is combine the individual classifier decision correctly to create the new combine model.

Bashir, S., [33] proposed HMV medical decision support framework using multi-layer classifier for disease prediction. They proposed based on the optimal combination of the heterogeneous classifier to create the ensemble model. The classifiers are Naïve Bayes, Linear Regression, and quadratic discriminate analysis, KNN, SVM, Decision Tree using Gini Index, and Decision Tree using Information Gain. So, their HMV ensemble framework outperforms the other prediction models. HMV framework was proposed based on three modules. The first module was data acquisition and preprocessing. The second module was used to predict unknown class label for test set instances. The third module used to predict and evaluate the proposed HMV ensemble model. After applying all the selected data set the HMV ensemble model achieved highest accuracy disease classification and prediction.

NaghmehKhajehali et al.[4] were presented by extracting factor affecting for pneumonia patients by using data mining techniques. They proposed modelling by using feature selection and classification with ensemble methods to preprocess, reduce dimensionality and classify the raw data. In their work, the design consists of different stages of preprocessing and used Bayesian Boosting method for constructed which identify factor related to patient LOS in hospital. The construction of modeling based on the data set SVM and ensemble method like AdaBoost, Vote, Stacking, Bayesian Boosting. Among these classifier techniques, Bayesian Boosting used for analysis of data by using 10 fold cross-validation method. In this work the data set was divided into 10 subsets, the training subset participated 10 times. Out of 10 subsets 9 were classified as the training set. The result indicated that the Bayesian Boosting ensemble technique was scored a better result. The Bayesian boosting ensemble technique, accuracy was 97.17%, which is high performance used to predict pneumonia disease in anticipation of LOS.

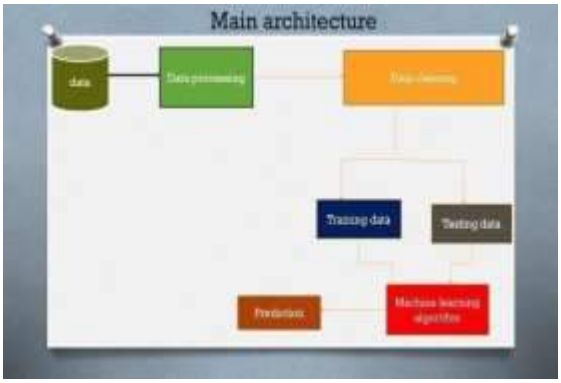
Pritom, A. I et al. [12] applied a classification algorithm for Predicting Breast Cancer Recurrence by using SVM, Decision tree, Naïve Bays and C4.5. They enhanced the accuracy of each classifier with the help of effective feature selection methods. They improve the accuracy by using Info Gain attribute with ranker search engine. After implemented on weka tool the recurrence prediction accuracy has SVM achieved 75.75% accuracy, J48 achieved 73.73% and naïve bays achieved 67.17%. These are the original data set without feature selection. After carefully applied feature selection SVM enhanced bay 1.52%, C4.5 enhanced by 2.52% and Naïve Bays enhanced by 9.09%.

Dulhare, U. N. et al. [10] Built classification models, Used feature selection to extract an action rule and predict CKD by using naïve Bayes classifier and one R attribute selector to predict and classify the CKD and none CKD patients. These methods are Naïve bays with the wrapper subset evaluator combine with the best first search. After implemented on weka tool using wrapper subset evaluators combine with the best first search engine; Naïve Bay's classifier achieved 97.5 % accuracy rate.

Architecture:

LOG and RF were selected as underlying components to generate the integrated model to improve the performance of judging. The probabilities that each sample was judged as notckd in LOG and RF were used as the outputs of underlying components. These two probabilities of each sample were obtained and could be expressed in a two-dimensional plane. In the complete CKD data sets, the probability distributions of the samples in a two-dimensional plane are similar. Therefore, the probability distribution of samples when K equalling to 11 is shown in Fig.3. It can be seen from Fig. 3 that the samples have different aggregation regions in the two-dimensional plane due to the different categories (ckd or notckd). In general, samples with ckd are concentrated in the lower left part, while the notckd samples are distributed in the top right part. Due to the fact that the results in the two models are different, some samples are located at the top left and lower right, and one of the two models makes the misjudgements. Perceptron can be used

to separate samples of two categories by plotting a decision line in the two-dimensional plane of the probability distribution. Ciaburro and Venkateswaran Defined perceptron as the basic building block of a neural network, and it can be understood as anything that requires multiple inputs and produces an output.



Objective Function: The perceptron used in this study is shown in Fig. 4. In Fig. 4, prob1 and prob2 are the probabilities that a sample was judged as notckd by LOG and RF,respectively. w0, w1 and w2 are the weights of input signals. w0 corresponds to 1, w1 corresponds to prob1 and w2 corresponds to prob2, respectively. y is calculated according to (7):

$y = w0 + w1 \times prob1 + w2 \times prob2.$

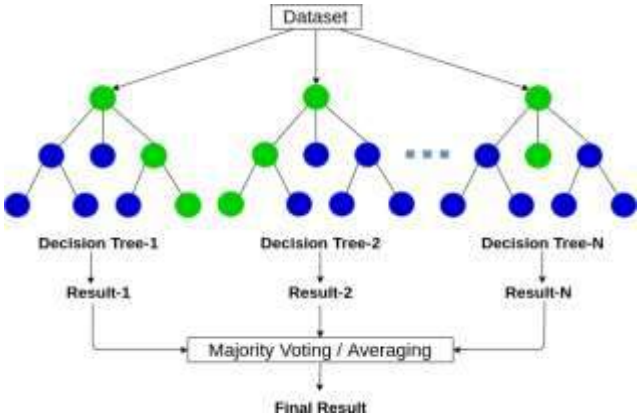
The probability distribution of the samples in the complete CKD data set (at K = 11) the horizontal axis and the vertical axis represent the probabilities that the samples were judged as not ckd by the LOG and the RF, respectively. Figure 4. The structure of the perceptron used in this study. The input signal corresponding to the weight w0 is 1, which is a bias. The function of Signum is used to calculate output by processing the value of y as follows: If $y > 0$, then the output = 1, whereas if $y < 0$, then the output = −1. For the output, 1 corresponds to not ckd, whereas -1 corresponds to ckd. A single perceptron is a linear classifier that can be used to detect binary targets. The weights are the core of the perceptron and adjusted in the training stage. $y = 0$ is the decision line, and this line can be described as(8):

prob2 = − w1 w2prob1 − w0 w2

ALGORITHM DESCRIPTION:

Randomforest:

- Random Forest is a trademark term for an ensemble of decisiontrees.
- In Random Forest, we've collection of decision trees (so known as"Forest").
- To classify an ew object based on attributes, each tree gives a classification nd we say the tree "votes" for that class.
- The forest chooses the classification having the most votes (over all the trees in the forest),



Logistic Regression:

- It is a classification not a regressionalgorithm.
- It is used to estimate discrete values based on given set of independentvariable(s).
- It predicts the probability of occurrence of an event by fitting data to a log it function. Hence, it is also known as logitregression.

Module Descriptions:

Data Collection Module:

Be it the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes.

Preparing the data Module:

Any analytical process thrives on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers. Exploratory analysis is perhaps one method to study the nuances of the data in details there by burgeoning the nutritional content.

Training a model:

This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test (proportion depending on the prerequisites); the first part (training data) is used for developing the model. The second part (test data), is used as a reference.

Disease prediction Module:

Disease will be predicted based on the results of the algorithms Provided.

Data Set:

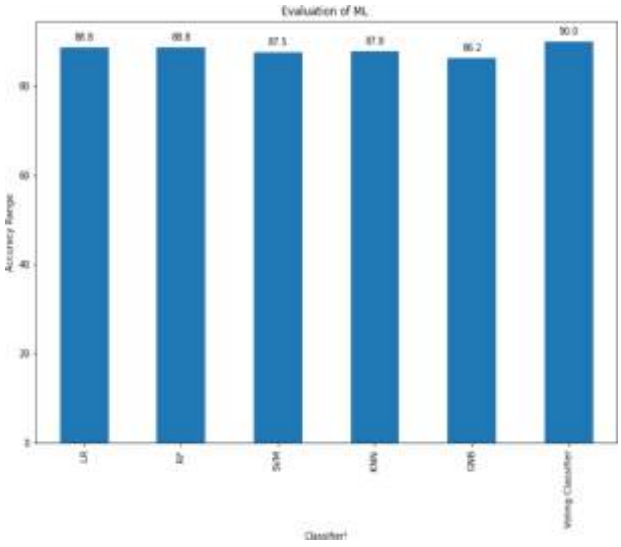
The CKD data set used in this study was obtained from the UCI machine learning repository.

TABLE 1. Details of each variable in the original CKD data set.

Variables	Explain	Class	Scale	Missing Rate
age	Age	Numerical	age in years	2.25%
bp	Blood Pressure	Numerical	in mm/Hg	3%
sg	Specific Gravity	Nominal	(1.005,1.010,1.015,1.020,1.025)	11.75%
al	Albumin	Nominal	(0,1,2,3,4,5)	11.5%
su	Sugar	Nominal	(0,1,2,3,4,5)	12.25%
rbc	Red Blood Cells	Nominal	(normal,abnormal)	38%
pc	Pus Cell	Nominal	(normal,abnormal)	16.25%
pcc	Pus Cell clumps	Nominal	(present,notpresent)	1%
ba	Bacteria	Nominal	(present,notpresent)	1%
bgr	Blood Glucose Random	Numerical	in mgs/dl	11%
bu	Blood Urea	Numerical	in mgs/dl	4.75%
sc	Serum Creatinine	Numerical	in mgs/dl	4.25%
sod	Sodium	Numerical	in mEq/L	21.75%
pot	Potassium	Numerical	in mEq/L	22%
hemo	Hemoglobin	Numerical	in gms	13%
pcv	Packed Cell Volume	Numerical	-	17.75%
wbcc	White Blood Cell Count	Numerical	in cells/cumm	26.5%
rbcc	Red Blood Cell Count	Numerical	in millions/cmm	32.75%
htn	Hypertension	Nominal	(yes,no)	0.5%
dm	Diabetes Mellitus	Nominal	(yes,no)	0.5%
cad	Coronary Artery Disease	Nominal	(yes,no)	0.5%
appet	appet	Nominal	(good,poor)	0.25%
pe	Pedal Edema	Nominal	(yes,no)	0.25%
ane	Anemia	Nominal	(yes,no)	0.25%
class	Class	Nominal	(ckd,notckd)	0%

Which was collected from hospital and donated by Sound a rapandian etal. on 3rd July, 2015. The data set contains 400 samples. In this CKD data set, each sample has 24 predictive variables or features (11 numerical variables and 13 categorical (nominal) variables) and a categorical response variable (class). Each class has two values, namely, Ckd (sample with CKD) and not ckd (sample without CKD). In the 400 samples, 250 samples belong to the category of ckd, whereas 150 samples belong to the category of not ckd. It is worth mentioning that there is a large number of missing values in the data. The details of each variable are listed in Table.

Evaluation of All Models:



Conclusion:

The proposed CKD diagnostic methodology is feasible in terms of data imputation and sample diagnosis. After unsupervised imputation of missing values in the data set by KNN imputation, the integrated model could achieve satisfactory accuracy. Hence, we speculate that applying this methodology to the practical diagnosis of CKD would achieve a desirable effect. In addition, this methodology might be applicable to the clinical data of the other diseases in actual medical diagnosis. However, in the process of establishing the model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples. Therefore, the

generalization performance of the model might be limited. In addition, due to there are only two categories (ckd and not ckd) of data samples in the data set, the model cannot diagnose the severity of CKD.

In the future, a large number of more complex and representative data will be collected to train the model to improve the generalization performance while enabling it to detect the severity of the disease. We believe that this model will be more and more perfect by the increase of size and quality of the data.

References:

1. K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, and S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques," CSI Trans. ICT, vol. 4, no. 2–4, pp. 313–319, Dec. 2016.
2. B. Boukenze, A. Haqiq, and H. Mousannif, "Predicting Chronic Kidney Failure Disease Using Data Mining Techniques," in Advances in Ubiquitous Networking 2, Springer, Singapore, 2017, pp. 701–712.
3. T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," Biomed. Signal Process. Control, Feb. 2017.
4. R. Ani, G. Sasi, U.R. Sankar, & O.S. Deepa, (2016, September) "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification - I EEE Conference Publication [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7732224/?reload=true. [Accessed: 15- Dec -2017].
5. L.-C. Cheng, Y.-H. Hu, and S.-H. Chiou, "Applying the Temporal Abstraction Technique to the Prediction of Chronic Kidney Disease Progression," J. Med. Syst., vol. 41, no. 5, p. 85, May 2017.
6. H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," J. Med. Syst., vol. 41, no. 4, p. 55, Apr. 2017.

7. P. Pangong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," in 2016 Second Asian Conference on Defence Technology (ACDT), 2016, pp. 145–150.
8. K.R.A. Padmanaban and G. Parthiban,"Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease," Indian J. Sci. Technol., vol. 9, no. 29, Aug. 2016.
9. S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," Procedia Comput. Sci., vol.82, no. Supplement C, pp. 115–121, Jan. 2016.
10. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naive Bayes classifier," in 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1–5.
11. N. Borisagar, D. Barad, and P. Raval, "Chronic Kidney Disease Prediction Using Back Propagation Neural Network Algorithm," in Proceedings of International Conference on Communication and Networks,Springer, Singapore, 2017, pp. 295–303.
A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," in 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 310–314.
12. S. Mishra, P. Chaudhury, B. K. Mishra, and H. K. Tripathy, "An Implementation of Feature Ranking Using Machine Learning Techniques for Diabetes Disease Prediction," in Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, New York, NY, USA, 2016, p. 42:1–42:3.
13. D. Zufferey, T. Hofer, J. Hennebert, M. Schumacher, R. Ingold, and S. Bromuri, "Performance comparison of multi-label learning algorithms on clinical data for chronic diseases," Comput. Biol Med., vol. 65, no. Supplement C, pp. 34–43, Oct. 2015.
14. S. Bashir, U. Qamar, F. H. Khan, and M. Y. Javed, "MV5: A Clinical Decision Support Framework for Heart Disease Prediction Using

Majority Vote Based Classifier Ensemble," Arab. J. Sci. Eng., vol. 39, no. 11, pp. 7771–7783, Nov. 2014.

15. T.R. Baitharu, & S.K. Pani, (2016). "Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder Dataset- Science Direct.[Online]Available: http://www.sciencedirect.com/science/article/pii/S1877050916306263. [Accesed: 15- Dec-2017].
16. Z. Sedighi, H. Ebrahimpour-Komleh, and S. J. Mousavirad, "Feature selection effects on kidney disease analysis," in 2015 International Congress on Technology, Communication and Knowledge (ICTCK), 2015, pp. 455–459.
17. D.M. Filimon, & A. Albu, (2014, May "Skin diseases diagnosis using artificial neural networks - IEEE Conference Publication." [Online]. Available: http://ieeexplore.ieee.org/abstract/document/6840059/. [Accessed: 15-Dec-2017]
18. F. Ahmad, N. A. M. Isa, Z. Hussain, M. K. Osman, and S. N. Sulaiman, "A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer," Pattern Anal. Appl., vol. 18, no. 4, pp. 861–870, Nov. 2015.
19. H. Asri, H. Mousannif, H. Al Moatassime, & T. Noel, (2016). "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis Science Direct." [Online].Available: http://www.sciencedirect.com/science/article/pii/S1877050916302575. [Accessed: 15- Dec- 2017].
20. B.L Deekshatulu, & P. Chandra, (2013 "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm ScienceDirect." [Online].Available: http://www.sciencedirect.com/science/article/pii/S2212017313004945. [Accessed: 15- Dec- 2017].
21. S. Ramya, and N. Radha. "Diagnosis of chronic kidney disease using machine learning algorithms."International Journal of Innovative Research in Computer and Communication Engineering vol. 4, no.1 pp. 812-820. 2016.

22. R. Dhruvi, R. Yavnika, & R. Nutan, " Prediction of Probability of Chronic Diseases and Providing Relative Real-Time Statistical Report using data mining and machine learning techniques". International Journal of Science, Engineering, and Technology Research (IJSETR) vol. 5, no. 4. 2016.
23. Compound feature generation and boosting model for cancer gene classification Ibrahim, S.Jafar Ali Ibrahim., Affir, A.M., Thangamani, M.International Journal of Engineering Trends and Technology, 2020, 68(10),pp.48–51,10.14445/22315381/IJETT-V68I10P208 https://ijettjournal.org/Volume-68/Issue-10/IJETT-V68I10P208.pdf
24. S. Vijayarani, S. Dhayanand, and M. Phil. "Kidney disease prediction using SVM and ANN algorithms. "International Journal of Computing and Business Research (IJCBR) vol. 6, no. 2, 2015.
25. N. Chetty, S. V. Kunwar, and S. D. Sudarsan. "Role of attributes selection in the classification of Chronic Kidney Disease patients "In Computing, Communication and Security (ICCCS), 2015 International Conference on, pp. 1-6. IEEE, 2015.
26. http://www.datasciencecentral.com/profiles/blogs/python-resources-fortop-data-mining-algorithms https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease
27. J. Radhakrishnan, and M. Sumit, "KI Reports and World Kidney Day." Kidney international reports vol.2, no. 2, pp. 125-126, Mar. 2017. https://www.kidney.org/kidneydisease/global-facts-about kidney disease#-ENREF-1
28. P. Ahmad, Q. Saqib, and S. Q. A. Rizvi. "Techniques of data mining in healthcare: a review." International Journal of Computer Applications Vol. 120, no. 15 Jan. 2015.
29. B.R.Sharma, K. Daljeet, and A.Manju. "Review on Data Mining: Its Challenges, Issues and Application."International Journal of Current Engineering and Technology vol. 3, no. 2 jun. 2013.
30. Bashir S., Qamar U., Khan F. H. & Naseem, L. (2016). H.MV: a medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science, 13, 10-25

31. Ibrahim, S. Jafar Ali, and M. Thangamani. "Prediction of Novel Drugs and Diseases for Hepatocellular Carcinoma Based on Multi-Source Simulated Annealing Based Random Walk." Journal of medical systems 42, no. 10 (2018): 188. https://doi.org/10.1007/s10916-018-1038-y

Detecting Counterfeit Banknotes With Machine Learning

Dr.R.Suneetha Rani [1], B.Gayathri [2], M.Venkata Surya [3], P.Jharani Asha Kiran [4], R.Siva Krishna [5]

Professor (IT), Dept. of Information Technology, QIS College of Engineering & Technology, India (1).

Final Year (B.Tech), Dept. of Information Technology, QIS College of Engineering & Technology, India (2,3,4,5)

Abstract:

The one important asset of our country is Bank currency and to create discrepancies of money miscreants introduce the fake notes which resembles to original note in thefinancial market. During demonetization time it is seen that so much of fake currency is floating in market. In general, by a human being, it is very difficult to identify forged note from the genuine not instead of various parameters designed for identification as many features of forged note are similar to original one. To discriminate between fake bank currency and original note is a challenging task. So, there must be an automated system that will be available in banks or in ATM machines. To design such an automated system there is need to design an efficient algorithm which is able to predict weather the banknote is genuine or forged bank currency as fake notes are designed with high precision. In this project six supervised machine learning algorithms are applied on dataset available on UCI machine learning repository for detection of Bank currency authentication

Introduction:

Financial activities are carrying out in every second by many persons in which one most important asset of our country is Banknotes. Fake notes are introduced in the market to create discrepancies in the financial market, even they resemble to the original note. Basically they are illegally created to complete various task. In 1990 forgery issue is not much of concern but as in late 19th century forgery has been increasing drastically. In 20th century

technology is increasing very vastly that will help the frauds to generate fake note whose resemblance is like genuine not and it is very difficult to discriminate them. This will lead to financial market to its lowest level. To stop this and to conduct smooth transaction circulation forged bank currency must be conserved. As a human being it is very difficult to identify between genuine and forged bank currency. Government have designed banknote with some features by which we can identify genuine. But frauds are creating fake note with almost same features with nice accuracy that make it very difficult to identify genuine note. So, now a days it is required that bank or ATM machines must have some system that can identify the forged note from the genuine note. To determine the legitimacy of the banknote artificial intelligence and Machine learning (ML) can play a vital role to design such a system that ca identify forged note from the genuine bank currency. Now a days, supervised machine learning (SML) approaches for classification problem is widely used. For medical disease its shows even promising results. Few authors have only applied SML algorithms on bank currency authentication. To identify weather a note is genuine or fake we have to develop an automation system. Initially, the input is an image of note and from different image processing techniques we can extract the features of note. Further these images are given as an input to the SML algorithms to predict whether note is original or fake. In review we can see that not much of work is done on this side.

Literature Survey:

2.1. Tushar Agasti, Gajanan Burand, Pratik Wade and P.Chitra-Fakecurrency detection using image processing:

Fake Currency has always been an issue which has created a lot ofproblems in the market. The increasing technological advancements have made the possibility for creating more counterfeit currency which are circulated in the market which reduces the overall economy of the country. There are machines present at banks and other commercial areas to check the authenticity of the currencies. But a common man does not have access to such systems and hence a need for a software to detect fake currency arises, which can be used by common people. This proposed system uses Image

Processing to detect whether the currency is genuine or counterfeit. The system is designed completely using Python programming language. It consists of the steps such as gray scale conversion, edge detection, segmentation, etc. which are performed using suitable methods.

Eshita Pilania, Bhavika Arora, Recognition of Fake Currency Based onSecurity Thread Feature of Currency:

In the last few years a great technological advances in color printing, duplicating and scanning, counterfeiting problems have become more serious. In past only authorized printing house has the ability to make currency paper, but now a days it is possible for anyone to print fake bank note with the help of modern technology such as computer, laser printer. Fake notes are burning questions in almost every country. Counterfeit notes are a problem of almost every country but India has been hit really hard and has become a very acute problem. Fake Indian currency of 100, 500 and 1000 rupees seems to have flooded the whole system and there is no proper way to deal with them for a common person. There is a need to design a system that is helpful in recognition of paper currency notes with fast speed and in less time. Our system describes an approach for verification of Indian and other countries currency banknotes. The currency will be verified by using image processing techniques.

2.3. Nayana Susan Jose, Shermin Siby, Juby Mathew, Mrudula Das, Android- Based Currency Recognition System for Blind:

In recent years, a lot of illegal counterfeiting rings manufacture and sell fake coins and at the same time fake note currency is printed as well which have caused great loss and damage to the society. Thus it is imperative to be able to detect fake currency We propose a new approach to detect fake Indian notes using their images. Currency image is represented in the dissimilarity space, which is a vector space constructed by comparing the image with a set of prototypes. Each dimension measures the dissimilarity between the image under consideration and a prototype. In order to obtain the dissimilarity between two images, the local key points on each image are detected and

described. Based on the characteristics of the currency, the matched key points between the two images can be identified in an efficient manner. A post processing procedure is further proposed to remove mismatched key points. Due to the limited number of fake currency in real life, SVM is conducted for fake currency detection, so only genuine currency are needed to train the classifier.

2.4. Komal Vora, Ami Shah, Jay Mehta, A Review Paper on Currency Recognition System:

In this paper, an algorithm based on the frequency domain feature extraction method is discussed for the detection of currency. This method efficiently utilizes the local spatial features in a currency image to recognize it. The entire system is pre-processed for the optimal and efficient implementation of two dimensional discrete wavelet transform (2D DWT) which is used to develop a currency recognition system. A set of coefficient statistical moments are then extracted from the approximate efficient matrix. The extracted features can be used for recognition, classification and retrieval of currency notes. The classification result will facilitate the recognition of fake currency mainly using serial number extraction by implementing OCR. It is found that the proposed method gives superior results.

PROPOSED SYSTEM:

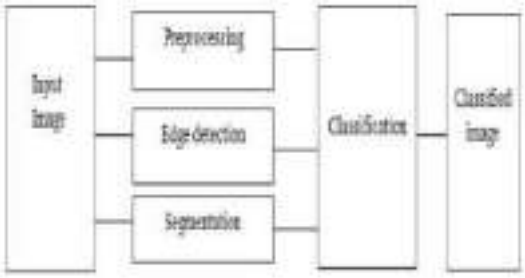
In proposed systems, we are using standard machine learning models like logistic regression, SVM, Navie bayes classification techniques. By company the accuracy of each model logistic regression gives better accuracy than other models. Logistic Regression is binary classification model. It follows probabilistic approach to classify currency note is false (or) original. We have visualized the dataset taken from UCI ML repository using different types of plotting, pre-processed data.

Architecture:

YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or

images. YOLO uses features learned by a deep convolutional neural network to detect an object. Versions 1-3 of YOLO were created by Joseph Redmon and Ali Farhadi. The first version of YOLO was created in 2016, and version 3, which is discussed extensively in this article, was made two years later in 2018. YOLOv3 is an improved version of YOLO and YOLOv2. YOLO is implemented using the Kera's or OpenCV deep learning libraries.

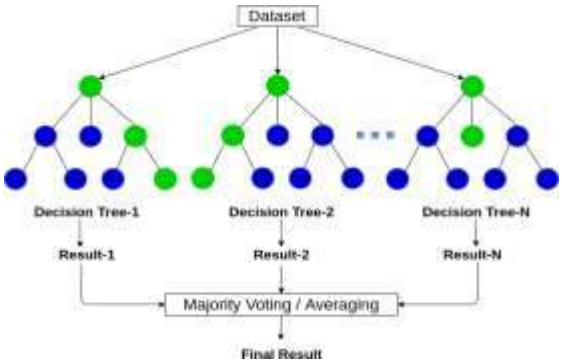
System architecture:



ALGORITHM DESCRIPTION:

Randomforest:

- Random Forest is a trademark term for an ensemble of decisiontrees.
- In Random Forest, we've collection of decision trees (so known as "Forest").
- To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class.
- The forest chooses the classification having the most votes (over all the trees in the forest).



Logistic Regression:

- It is a classification not a regressionalgorithm.
- It is used to estimate discrete values based on given set of independentvariable(s).
- It predicts the probability of occurrence of an event by fitting data to a log it function. Hence, it is also known as logitregression.

Module Descriptions:

Dataset: The dataset used is Indian currency. The dataset contains various Indian currencies, euro's, dollars.

Upload image: In this module user can upload currency image, based on input image it will show type of currency note and percentage of note.

Classification: Classification represents the matter of distinctive to that of a group of classes a new observation belongs, on the idea of a training set of data having observations whose category membership is known.

Test & Train Set: A training dataset could be a dataset of eg's used for learning, that is to fit the parameters for eg, a classifier. A test dataset could be a dataset that is independent of the training dataset, but follows a similar likelihood distribution as the training dataset. If a model fit to the training dataset conjointly fits the test dataset well, a lowest over fitting takes place.

Prediction: We compare the given input image with trained data (Yolov3) in CNN algorithm by using Random forest classifier to detect the fake percentage of currency.

Evaluation of All Models:

Conclusion:

Hand Gesture recognition and voice conversion for dumb and deaf person was successfully executed using image processing. The method takes image as input and gives text and speech as an output. Implementation of this system gives up to 90% accuracy and works successfully in most of the test cases.

References:

1. M. Aoba, T. Kikuchi, and Y. Takefuji, "Euro Banknote Recognition System Using a Three-layered Perceptron and RBF Networks", IPSJ Transactions on Mathematical Modeling and it's Applications, May 2003.
2. S. Desai, S. Kabade, A. Bakshi, A. Gunjal, M. Yeole, "Implementation of Multiple Kernel Support Vector Machine for Automatic Recognition and Classification of Counterfeit Notes", International Journal of Scientific & Engineering Research, October-2014.
3. C.Gigliaranoa, S.Figini, P. Muliere, "Making classifier performance comparisons when ROC curves intersect", Computational Statistics and Data Analysis 77 (2014) 300–312.
4. E. Gillich and V. Lohweg, "Banknote Authentication", 2014.
5. H Hassanpour and E.Hallajian, "Using Hidden Markov Models for Feature Extraction in Paper Currency Recognition.
6. Z. Huang, H. Chen, C. J. Hsu, W. H. Chen and S. Wuc, "Credit rating analysis with support vector machines and neural network: a market comparative study", 2004.
7. C. Kumar and A. K. Dudyala, "Banknote Authentication using Decision Tree rules and Machine Learning Techniques", International Conference on Advances in Computer Engineering and Applications(ICACEA), 2015.
8. M. Lee and T. Chang, "Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress", International Journal of Artificial Intelligence & Applications 1.3 (2010) 31-43.
9. C. Nastoulis, A. Leros, and N. Bardis, "Banknote Recognition Based On Probabilistic Neural Network Models", Proceedings of the 10th WSEAS International Conference on SYSTEMS, Vouliagmeni, Athens, Greece, July 10-12, 2006.
10. S. Omatu, M. Yoshioka and Y. Kosaka, "Bankcurrency Classification Using Neural Networks", IEEE, 2007.
11. A. Patle and D. S. Chouhan, "SVM Kernel Functions for Classification", ICATE 2013.

12. E. L. Prime and D. H. Solomon, "Australia's plastic banknotes: fighting counterfeit currency.," Angewandte Chemie (International ed. in English), vol. 49, no. 22, pp. 3726–36, May 2010.
13. A. Roy, B. Halder, and U. Garain, "Authentication of currency notes through printing technique verification," Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP '10, pp. 383–390, 2010.
14. P. D. Shahare and R. N. Giri, "Comparative Analysis of Artificial Neural Network and Support Vector Machine Classification for Breast Cancer Detection", International Research Journal of Engineering and Technology, Dec-2015.
15. F. Takeda, L. Sakoobunthu and H. Satou, "Thai Banknote Recognition Using Neural Network and Continues Learning by DSP Unit", International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 2003.
16. M. Thirunavukkarasu, K. Dinakaran, E.N Satishkumar and S Gnanendra, "Comparison of support vector machine(svm) and Back propagation network (bpn) methods in predicting the protein Virulence factors"J of Industrial Polluction Control 33(2)(2017)pp 11-19.
17. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International Multi Conference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352357http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf

Image Caption Generator Using CNN and LSTM

Dr. G. Lakshmi Vara Prasad[1], B. Sri Mounika[2], P. Vijaybabu[3], A. Teethes babu[4],Ch.Srikanth[5]

[1]Associate Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India 2,3,4,5, Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Abstract - For this paper, we use CNN and LSTM to become aware of the caption of the image. Image caption generation is a system that comprehends natural language processing & computer vision standards to recognize the connection of the image in English. In this research paper, we cautiously pursue a number of important concepts of photograph captioning and its familiar processes. We talk about Kera's library, NumPy and Jupiter notebooks for the making of this paper. We also talk about Flickr dataset and CNN used for photo classification.

Keywords- CNN, LSTM, image captioning, deep learning.

1. INTRODUCTION:

Every day we see a lot of photographs in the surroundings, on social media and in the newspapers. Humans are able to recognize photographs themselves only. We humans can pick out the photographs without their designated captions but on the other hand machines need images to get trained first then it'd generate the photograph caption automatically. Image captioning may benefit for loads of purposes, for example supporting the visionless person using text-to-speech through real time feedback about encompassing the situation over a camera feed, improving social medical leisure with the aid of reorganizing the captions for photographs in social feed alongwith messages to speech. Facilitating kids in recognizing substances further to gaining knowledge of the language. Captions for every photograph on the world wide web can produce quicker & detailed authentic photographs exploring and indexing. Image captioning has diverse packages in numerous fields

inclusive of biomedicine, commerce, internet looking and navy and many others. Social media like Instagram, Facebook etc. can generate captions routinely from images The principal goal of this research paper is to get a little bit of expertise in deep learning strategies. We use two strategies specially CNN and LSTM for image classification.

2. BACKGROUND WORK:

IMAGE CAPTIONING TECHNIQUES :

CNN - Convolutional Neural systems are specific important neural systems that can produce information that has an information shape, for example, a 2D lattice and CNN is valuable for working with pictures. It examines pictures from left corner to the right corner and through to extricate significant highlights from the picture, and consolidates the element to characterize pictures. It can deal with interpreted, pivoted, scaled, and modified pictures. The Convolutional neural system is a profound learning calculation that takes in the info picture, allocates significance to various components/protests in the picture, and recognizes it from each other.

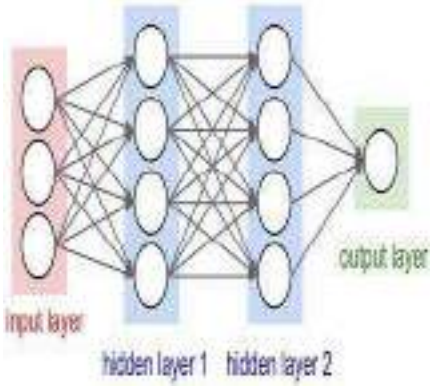


Fig. Regular Neural Network

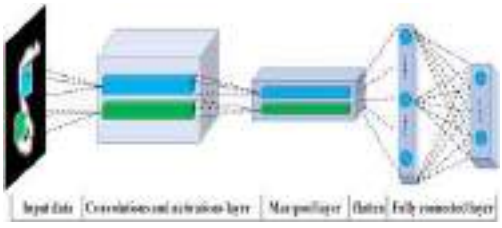


Fig. Convolutional Neural Network

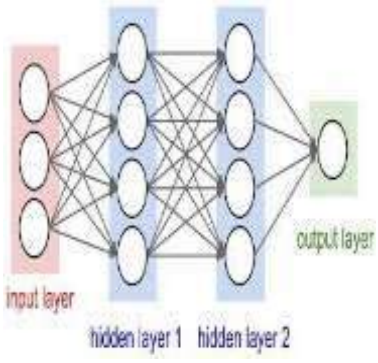
Fig.1 CNN Architecture:

The required pre-handling in ConvNent negligible when compared with other order calculations. In spite of the fact that channels are hand-designed in crude strategies, with sufficient preparation, ConvNets is fit for learning these channels/highlights. The structure of the curved system is like the neuronal network design inside human mind & is inspired by the way of the organization of the visual cortex. Singular neurons react to upgrades simply in a limited district of the visible field known as open field. The assortment of such fields covers the summation of visual regions.

CNN : Architecture - A pure rustic neural network, in whatever location all neurons in a single layer merge with all of the neurons in the subsequent layer is inefficient in regards to analyzing large pictures and video. For a normal size picture with many picture elements called pixels & 3-tone colors (RGB i.e., red color, greencolor, blue color), the range of restriction utilizing an accepted neural system will be in the tons, & that can prompt overfitting.

To constrain effective quantities of restrictions & recognition of the neural system on significant pieces of picture, CNN utilizes a 3D arrangement in which each adjustment of neurons breaks down a little area or "highlight" of picture. Rather than all neurons to skip their selections to the next neural layer, each gathering of the neurons spends significant time in distinguishing one piece of picture, such as a nose, a left ear, mouth or a leg. The last yield is a point of scope, illustrating how reasonable every one of abilities is elected as part of the class.

A Traditional neural network structure (left)



3 -Dimensional CNN structure (right):

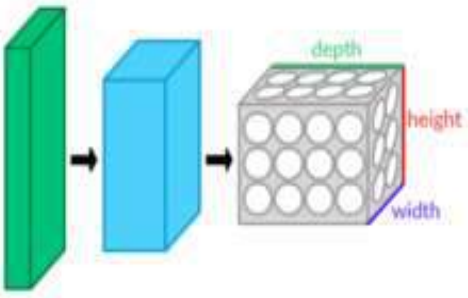
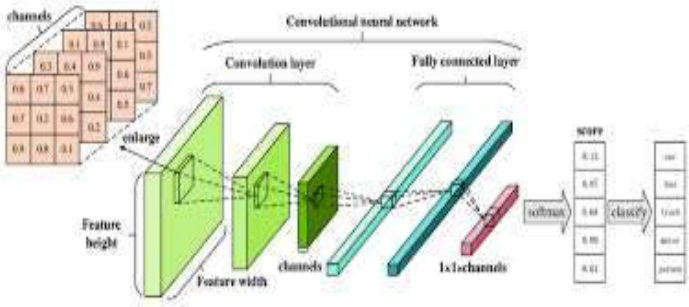


Fig.2 Working of CNN

How does CNN Work :

As we have discussed previously, a fully connected neural network where the input in the preceding layers is connected to every input in the following layers is convenient for the task at hand, along such lines, according to CNN, the neurons in a cell may be connected with a specific cell area before it, rather



This helps in reducing the complexity of the neural network and acquiring less computing power. As per new computer under standard image with the use of numbers at each pixel. When we generally compare two images we check the pixel values of each pixel. This technique only helps us to compare two identical images only but when we keep different images to compare the comparison fails. In CNN image comparison takes place piece by piece.

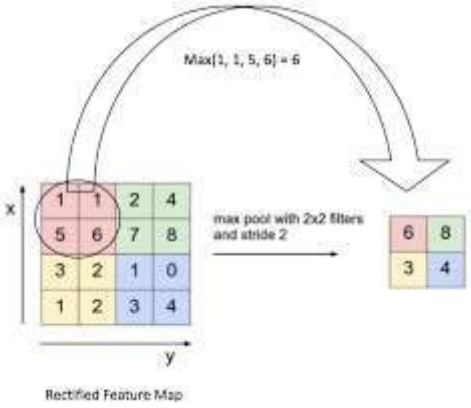


Fig.3 Feature map of CNN picture

The main reason behind using CNN algorithm is that, this is the only algorithm which takes pictures as an input and on the basis of input pictures drawing the feature map, ie. classifying each pixel on the basis on similarity and differences. The CNN classifies the pixels and a matrix is created, which is known as feature map. Feature map is a collection of similar pixels placed in a separate category. These matrices play an important role in finding the essence of the thing in the input picture.

More about CNN:

There are total 3 types of layers in CNN model

1. Convolutional
2. Pooling
3. Fully connected

In the first layer, the input image is read through the CNN, and on that foundation a feature map is made. From that feature map, it serves as an input to the following layers, i.e. for the Pooling layer. In the pooling layer, the feature map is broken down into extra simpler parts to carefully examine the context of the picture. This layer makes the feature map more dense so as to discover the most critical information about the picture.

The 1st and 2nd layers i.e., Convolutional and Pooling they're practiced so many times, depending on the picture as to get the densed information about the picture. The extra dense feature map is created because

of these two layers. And this densed feature map is utilized by the last layer i.e., Fully Connected. This layer performs classification. It sorts the pixels with respect to similarity and differences. Classification is done up to exceptional limit so as to get the essence of the picture, help objects, persons, things, etc.

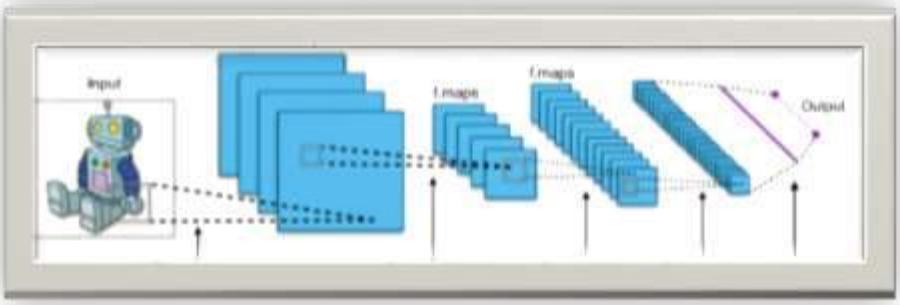


Fig.4 Layers of scanned picture

These layers help CNN to clearly locate and find features of the picture. Extraction of vital features present in the picture of fixed length inputs is transformed into fixed size outputs.

CNN techniques are very much in usage viz,

Computer vision: in the area of medical sciences image analysis is done through CNNs only. Inner structure of the body is effortlessly examined with the help of this.

In mobile phones, it's been used for so many things, for instance, to find the age of the person, to unlock the phone by examining the picture from the camera.

In industries its far used for making patents or copyright of specific clicked pictures.

Pharmaceuticals discovery: Its been broadly used for discovering the drugs/pharmaceuticals, by analyzing the chemical features and finding the best drug to cure a particular problem.

Origin of LSTM: LSTM was first searched by two German researchers - Sepp Hochester and Jurgen Schmid Huber, in 1997. LSTM stands for long

short-term memory. In the Deep Learning discipline of recurrent neural networks, LSTM holds a crucial place. The special element about LSTM is that it not only stores the input data, but can also supply predictions about the subsequent datasets through its own. This LSTM network retains the stored data for a particular time period and on that basis predicts or gives the future values to the data. This is the main purpose why LSTM is used here more than that of traditional RNN.

The Problem with RNNS (Recurrent Neural Networks):

RNNs are a part of a deep learning set of rules which are performed to deal with a number of complicated or complex computer tasks like item classification & speech recognition. RNNs are performed to address an array of activities that arise in series, with the information of every situation based completely on statistics from preceding situations.

Exquisitely, we intend to favor RNNs which are having extended collections of data & higher capabilities. This RNN can be used to carry out plenty of real life problems like inventory forecasting & reinforce speech recognition. Yet, RNNs are not used to solve real life problems & that is because of the Vanishing Gradient problem.

Vanishing Gradient Problem:

This vanishing gradient problem is the main cause which makes the working of RNNs challenging. In general, the engineering of RNNs is made such that it stores the data for some short period of time and stores some array of data. It's not possible for RNNs to remember all the data values and a long period of time. RNNs can only store some of the data for a small period of time. Thereupon, the reminiscence of RNNs is only favorable for shorter arrays of data and for short-time periods. This vanishing gradient problem becomes very prominent as compared to traditional RNNs- to solve a particular problem it adds so many time steps, which results in losing the data when we use backpropagation. With so many time steps, RNNs have to store data values of each time step, which results in storing more & more data values and that one is not feasible in the case of RNNs. And by this vanishing gradient problem is formed.

What can be done so as to solve this Vanishing Gradient problem with RNNs :

To solve this problem, we will be using Long short-term memory (LSTM) , which is a subset of RNNs. LSTM are basically constructed to overcome the problem of Vanishing Gradients. The exceptional thing about LSTM is that it can preserve the data values for lengthy interval of time and hence can solve the vanishing gradient problem LSTMs are constructed in such a manner that they always contain errors. And due to these errors LSTM keeps studying the data values over several time steps. Because of studying data values again & again, it makes studying easy over time & layers.

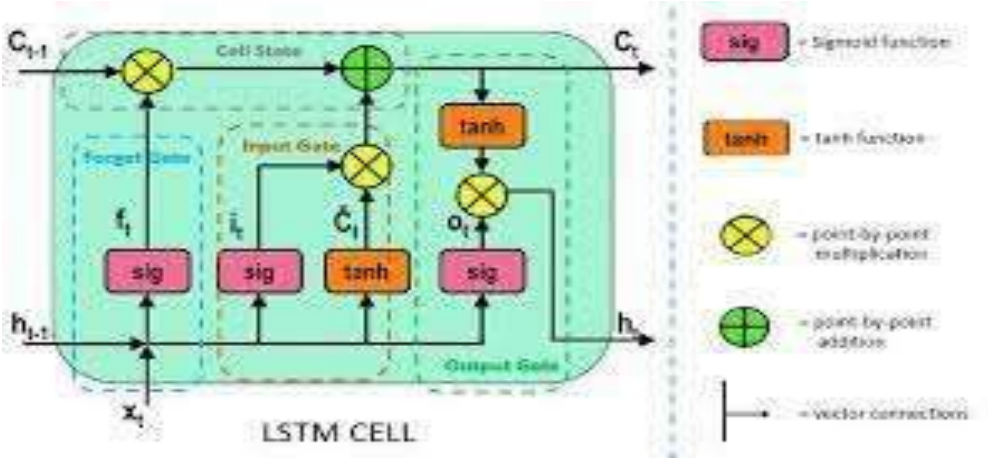


Fig.5 Gates in LSTM

Ideally, as per the diagram above, LSTM uses several gates to store the data and after that it processes the data and sends the result to the final gate. When we talk about RNNs, they used to pass the data to the final gate without any processing. From these gates in LSTM, the whole network can shape the data in many forms, including storing the data and reviewing the data from the gates. The gates in the LSTM are independently skillful to make judgement concerning the facts & the data. Moreover, these gates are able to make judgements on their own by opening or closing the gates.

The understanding of LSTM gates to hold the data for a period of time offers benefit to the LSTM over the RNNs.

Architecture Of LSTM :

The architecture of LSTM is very simple, it consists of 3 major gates, which store the data for a longer period of time and help in solving the difficulties which RNNs couldn't solve.

The 3 major gates of the LSTM covers are: **Forget gate** — the main work of the forget gate is to filter the data, i.e. to delete all that data which is not needed in the future to solve a particular task. This gate is responsible for the overall performance of the LSTM, it optimizes the data.

Input gate: the starting of LSTM starts from this gate, i.e. input gate. This gate takes input from the user and supplies the input data to other gates. ·

Output gate: This gate is responsible for showcasing the desired result in a proper manner.

Uses of Long Short-Term Memory Networks : LSTMs are profoundly and mostly used for variety deep learning duties that largely encompasses forecasting of the data depending upon the preceding data. The 2 remarkable illustrations cover text prediction and stock market prediction.

Text Prediction - The LSTM is very used in predicting the texts. The long term memory, understanding of LSTM makes it capable enough to predict the next words in the sentences. This is the result of the LSTM network in predicting the next words by its own. The LSTM first stores the data, the feel of the words, the styling of the words, the use of the words in a particular situation,etc and on that basis predicts the next words. The stored data, i.e. input data is further used for future use.

The best illustration can be given of text prediction is a Chatbot, that is widely utilized by the eCommerce websites and mobile applications.

Stock market Prediction - In the stock market also, LSTM stores the data or the trends in which the market behaves at a particular time, at a particular instant and on that account predicts the next variations and trends of the

market. It's a problematic task to predict the variation in the stock market because market variations are very challenging to predict and forecast. The LSTM model has to be trained in such a manner that it gives the correct values to the users. For that, a lot of data has to be stored for a lot of time, it can take days also.

More about LSTM - LSTMs are basically a part of RNNs, which are having capacity to hold more data values as compared to RNNs. LSTMs are widely in use today in every field. The simplest diagram of LSTM is shown below. It consists of 3 major gates viz, Forget gate, input gate, output gate. These gates are having capacity to store the data and give out the desired output. Whenever talked about LSTM network the three gates always comes up. The below diagram shows the simplest architecture of LSTM :

3. PROPOSED METHOD:

Image Caption Generation Model:

In order to prepare an image caption generation model, we will be summing up the two different architectures. It is further called as CNN-LSTM model. So, in this we will be using these two architectures to get the caption for the input pictures.

CNN - it's been used to extract the important features from the input picture. To do this, we have taken a pre-trained model for our consideration named Xception.

LSTM -it's been used to store the data or the features from the CNN model and further process it and to support in the generation of a good caption for the picture.

Project File Architecture: For our research purpose, we have downloaded the data set which consists of following files:

Flickr8k-Data sets : This file contains all the pictures for which we have to first train our model. It contains 8091 images.

Flickr8k- texts : This folder contains text files & pre-formed captions for the pictures. The following files are set up for making this system to run by us to check the working of the CNN-LSTM model.

Model - This folder will contain all the trained models which are at first trained. This would be one time process to train the model.

Description.txt - This is the file which will contain the picture names & their related captions later preprocessing.

Feature.p - This file binds the picture and their related captions that are extracted from the Xception, which is a pre-trained CNN model.

Tokenizers.P - This file contains an expression which we call tokens , and these tokens are generalized with the index value.
Models.png – Diagrammatic representation of extension of the CNN-LSTM model.

Testing- captions- generator.py : This is the Python file which is used in generating the captions of the pictures.

Training- captions- generator. Ipynb : This is basically a Jupiter notebook, which is in short, a web-based application. We use this to train our model & on that basis achieving captions to our input pictures.

4. CONCLUSION:

The CNN-LSTM model was built on the idea of generating the captions for the input pictures. This model can be used for a variety of applications. In this, we studied about the CNN model, RNN models, LSTM models, and in the end we validated that the model is generating captions for the input pictures.

REFERENCES :

1. Abhaya Agarwal and Alon Lavie. 2008. Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. In Proceedings of the Third Workshop on Statistical Machine Translation. Association for Computational Linguistics, 115–118.
2. Ahmet Aker and Robert Gaizauskas. 2010. Generating image descriptions using dependency relational patterns. In Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, 1250–1258.
3. Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. 2016. Spice: Semantic propositional image caption evaluation. In European Conference on Computer Vision. Springer, 382–398.
4. Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2017. Bottom-up and top-down attention for image captioning and vqa. arXiv preprint arXiv:1707.07998 (2017).
5. Jyoti Aneja, Aditya Deshpande, and Alexander G Schwing. 2018. Convolutional image captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5561–5570.
6. Lisa Anne Hendricks, Subhashini Venugopalan, Marcus Rohrbach, Raymond Mooney, Kate Saenko, Trevor Darrell, Junhua Mao, Jonathan Huang, Alexander Toshev, Oana Camburu, et al. 2016. Deep compositional captioning: Describing novel object categories without paired training data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
7. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations (ICLR).
8. Shuang Bai and Shan An. 2018. A Survey on Automatic Image Caption Generation. Neurocomputing. ACM Computing Surveys, Vol. 0, No. 0, Article 0. Acceptance Date: October 2018. 0:30 Hossain et al.

9. Satanjeeve Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, Vol. 29. 65–72
10. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International Multi Conference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352357http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf
11. Jafar Ali Ibrahim. S, Mohamed Affir. A "Effective Scheduling of Jobs Using Reallocation of Resources Along With Best Fit Strategy and Priority", International Journal of Science Engineering and Advanced Technology (IJSEAT) – ISSN No: 2321- 6905, Vol.2, Issue.2, Feb-2014, http://www.ijseat.com/index.php/ijseat/article/view/62

Crime Data Anlysis Using Machine Learning

P.Kiran[1], B.Mounika[2], P.Naveen[3], N.Tejaswini[4], B.Karthik[5]

Assistant Professor (IT), Dept of Information Technology QIS College of Engineering & Technology, Andhra Pradesh, India (1)
Final Year (B.Tech), Dept. Of Information Technology, QIS College Eng Andhra Pradesh,India(2,3,4,5)

ABSTRACT: Crime is one of the biggest and dominating problems in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, Machine Learning and data science techniques are used for crime prediction of Chicago crime data set. For this supervised classification Random Forest algorithm is used. This approach involves predicting crimes classifying, pattern detection and visualization with effective tools and technologies. Use of past crime data trends helps us to correlate factors which might help understanding the future scope of crimes.In this work, various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets are processed and visualized based on the need.

KEYWORDS: crime analysis, prediction analysis, Machine Learning, decision trees, pattern detection.

I. INTRODUCTION:

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and

spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

Latest technical developments in sophisticated tools of dataanalytics and visualization are helping the society in differentways to analyze the data of social relevance. One of suchsocially relevant activities is crime details of differentdemographic places. The analysis of the crime data will helpdecision making agencies to take precautionary steps to controlthe crime rate over demographic places.

Today, a high number of crimes are causing a lot of problems in many different countries. In fact, scientists are spending time studying crime and criminal behaviors in order to understand the characteristics of crime and to discover crime patterns. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular, issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis. Dealing with crime data is very challenging as the size of crime data grows very fast, so it can cause storage and analysis problems. In particular, issues arise as to how to choose accurate techniques for analyzing data due to the inconsistency and inadequacy of these kinds of data. These issues motivate scientists to conduct research on these kinds of data to enhance crime data analysis. Data mining and machine learning are inter-disciplinary fields involving computers and mathematics wherein the programming is done for the system to carry out the operation. Both are highly important in detection and prevention of crime. Crime analysis involves extraction of crime patterns, prediction and crime detection.

The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. This work helps the law enforcement agencies to predict and detect crimes in Chicago with improved accuracy and thus reduces the crime rate. There has been tremendous increase in machine learning algorithms that have made crime prediction feasible based on past data. The aim of this project is to perform analysis and prediction of crimes in states using machine learning models. It focuses on creating a model that can help to detect the number of crimes by its type in a particular state.

The rest of the paper is organized as follows: Section II demonstrates the literature survey of different crime prediction models. The section III presents crime prediction and analysis using ML classifiers. The result analysis is demonstrated in section IV finally conclusion is presented in section V.

II. LITERATURE SURVEY:

Mohammed Boukabous, Mostafa Azizi et. al., [1] presented a "Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers (BERT).In this paper, a hybrid approach is used that combines both lexicon-based and deep learning, with BERT as the DL model. The authors employed the lexicon-based approach to label our Twitter dataset with a set of normal and crime-related lexicons; then, we used the obtained labeled dataset to train this BERT model.

Ricardo Francisco ReierForradellas et. al., [2] Applied Machine Learning in Social Sciences: Neural Networksand Crime Prediction.This study proposes a crime prediction model according to communes. For this, the Python programming language is used,due to its versatility and wide availability of libraries oriented to Machine Learning. For prediction, it is necessary to provide the modelwith the information corresponding to the predictive characteristics (predict); these characteristicsbeing according to the developed neural network model: year, month, day, time zone, commune,and type of crime.

Neil Shah, Nandish Bhagat etal., [3] presented a crime forecasting approach using ML and computer vision methods for the prediction and prevention of crime.In this paper, they described the results of certain cases where such approacheswere used, and which are motivated to pursue further research in this field The sole purpose of this study is to determine how a combination of ML and computer vision can beused by law agencies or authorities to detect, prevent, and solve crimes at a much more accurate and faster rate.

Panagiotis Stalidis etal., [4] Examining Deep Learning Architectures for Crime Classificationand Prediction. In this paper, a detailed study on crime classification and prediction using deep learning architectures is presented. The author examine the effectiveness of deep learning algorithms in this domainand provide recommendations for designing and training deep learning systems for predictingcrime areas, using open data from police reports.

PawełCichosz, etal., [5] has discussed about the Urban Crime Risk Prediction Using Point ofInterest (POI) Data.This articledemonstrates how they can be combined with ML algorithms to create crime predictionmodels for urban areas. Selected POI layers from OpenStreetMap are used toderive attributes describing micro-areas, which are -3assigned crime risk classes based on police crimerecords.

Shaobing Wu et. al., [6] presented Crime Prediction Using Data Miningand Machine Learning.The aim of the study is to show the pattern and rate of crime based on the data collected and to show the relationshipsthat exist among the various crime types and crime Variable.Byintroducing formula and methods of Bayesian network, random tree and neuralnetwork in machine learning and big data, to analyze the crime rules from thecollected data.

Sarah Brayne et. al., [7] presentedTechnologies of Crime Prediction: TheReception of Algorithms in Policing andCriminal Courts. We draw on ethnographic fieldwork conducted within a large urban police departmentand a midsized criminal court to assess the impact of predictive technologies atdifferent stages of the criminal justice process. The authorstudied how predictive algorithms are used, documentingsimilar processes of professional resistance among law enforcement and legal professionals.

Shraddha Ramdas Bandekar et. al., [8] presentedDesign and analysis of Machine Learning Algorithms for the reduction of crime rates in India. This research work focuses on how machine learning algorithms can be designed and analyzed to reduce crime rates in India. By the means of machine learning techniques, determining the pattern relations among huge set of data has become easier. This research mainly depends on providing a prediction on crime type that might occur based on the location where it has already taken place. Machine learning has been used to develop a model by the use of training data set that have gone through the process of data cleaning and transformation.

Chao Huang, Junbo Zhang et. al., [9] presented DeepCrime: Attentive Hierarchical Recurrent Networks forCrime Prediction.In this paper, we develop a new crime predictionframework–*DeepCrime*, a deep neural network architecture thatuncovers dynamic crime patterns and carefully explores the evolvinginter-dependencies between crimes and other ubiquitous datain urban space. Extensive experiments on real-world datasets demonstrate the superiority of this framework over many competitivebaselines across various settings.

Shubham Agarwa et. al., [10] presented Crime Prediction based on Statistical Models.Based on the previous year (s) crime details in Indian states, the author present statistical models *viz.*Weighted Moving Average, Functional Coefficient Regressionand Arithmetic-Geometric Progression based prediction of thecrime in coming years.

III. CRIME PREDICTION AND ANALYSIS USING ML MODELS:

This data set includes criminal offenses in the City and County for the previous five calendar years plus the current year to date. The data is based on the National Incident Based Reporting System (NIBRS) which includes all victims of person crimes and all crimes within an incident.Machine learning agents work with data and employ different techniques to find patterns in data making it very useful for predictive analysis.

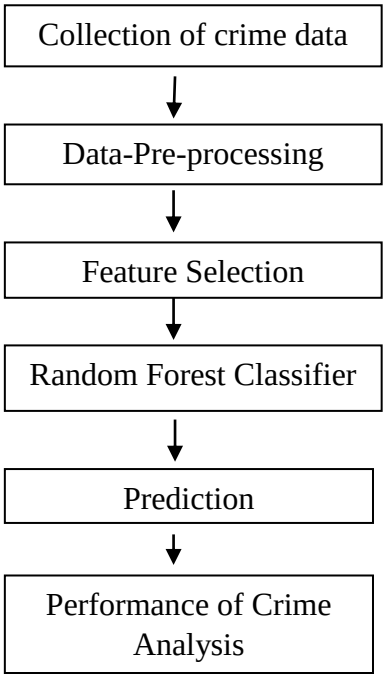


Fig 1: FLOW CHART OF PRESENTED CRIME PREDICTION AND ANALYSIS MODEL

This data reflects reported incidents of crime that have occurred in the City of Chicago during a specific time period. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. This data set contains 90 days of information and the most recent data available is seven days prior to today's date.Here a Data set is analyzed by using a object based approach with in the Python programming Language.Building the model will be done using better algorithm depending upon the accuracy.

Data processing is generally the collection and manipulation of itemsof data to produce meaningful information. In this sense it can beconsidered a subset of information processing, the change ofinformation in any manner detectable by an observer.This simple observation led to the idea that it would be useful to useonly some selected trees in classification. The selection of trees wasbased on their performance on similar instances, but without success.The

step toward the analysis is preprocessing. If the data is dirty, it willgenerate incorrect visualizations, hence leading to wards the incorrect conclusions. The crime data collected also has some level of dirtiness.

It contains some null values, inconsistent date formats, and someoutliers.

Feature selection is also known as variable selection. It is the automaticselection of attributes in data that are most relevant to the predictivemodeling problem.Random split selection does better than bagging; introduction of random noise into the outputs also does better; but noneof these do as well as adaboost by adaptive reweighting (arcing) of thetraining set.

The importance of each feature variable in a trainingsubset refers to the portion of the gain ratio of the variable comparedwith the total feature variables. The value of all feature variables aresorted in descending order and the top variable values are selected.Thus the number of dimensions of the dataset is reduced from featurevariables in each sample to the number of the selected featurevariables.

To make a prediction on a new instance, a random forest mustaggregate the predictions from its set of decision trees. Thisaggregation is done differently for classification and regression.Thedata gets split into many subsets and it compares the train and test datato and the best one. This process gets repeatedly on each subset andfind out the best prediction on each mapping. According to thisprocess, each subset has its own predicted class. And comparing all thepredicted class of its produce the final prediction based on trainingdata.

Here in the proposed system we use the random forest algorithm in order to get good results and better accuracy when compared to the above or the existing algorithms. We use random forest for accuracy. Random forest is a most popular and powerful supervised machine learning algorithm capable of performing both classification, regression tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Random decision forests correct for decision trees habit of over fitting to their training set .The data sets considered are rainfall, perception, production, temperature to construct random forest, a collection of decision trees by considering two-third of the records in the datasets. These decision trees areapplied on the remaining records for accurate classification.

IV. RESULT ANALYSIS:

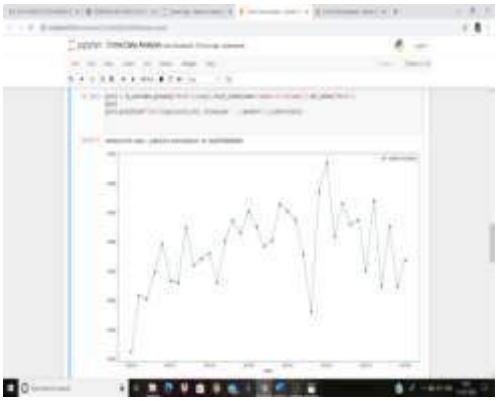
The presented system analysis is analyzed by following steps. Python programming Language provides libraries which also focus on the implementation of data analysis and in representing the data in different visualizations.When you install Python directly from its official website, it does not include Jupyter Notebook in its standard library. In this case, you need to install Jupyter Notebook using the pip. The process is as follows: Open a new command prompt (Windows) or terminal (Mac/Linux). Execute the following command to install Jupyter Notebook.

Fig. 2: Python3 -m Pip Install Jupyter Window

After you have installed the Jupyter Notebook on your computer, you are ready to run the notebook server.Keep the terminal open as it is. It will then open the default web browser with the URL mentioned in the command prompt or terminal. When the notebook opens in your browser, you will see the Notebook Homepage as shown. This will list the notebook files and subdirectories in the directory where the notebook server was started.After executing the code, we get the resultant output.The following screenshots shows the resultant output.

Fig. 3: SCREENSHOTS OF RESULTANT OUTPUTS

The Fig. 4 represents the crimes per month.



The Fig. 5 represents the bar graph of crime wise arrests.

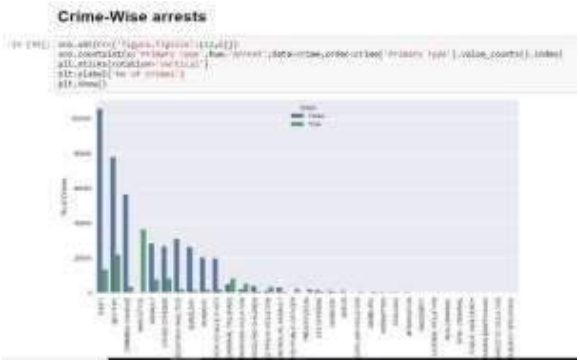


Fig. 5: CRIME WISE ARRESTS BAR GRAPH

The Fig. 6 represents the number of crimes occurred per an hour i.e. HEAT MAT of crime.

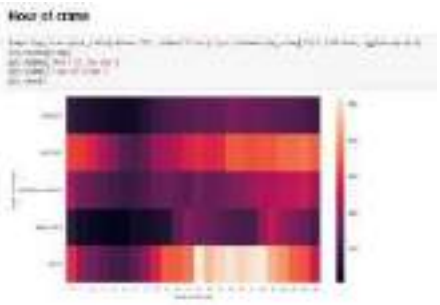


Fig. 6: CRIMES HEAT MAP (HOUR OF CRIME OCCURRED)

V. CONCLUSION:

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with accuracy of 0.789. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding Chicago crimes datasets that can help in capturing the factors that can help in keeping society safe. The tool we have developed provides a framework for visualizing the crime networks and analyzing them by various machine learning algorithms using the Google Maps. The project helps the crime analysts to analyze these crime networks by means of various interactive visualizations. The interactive and visual feature applications will be helpful in reporting and discovering the crime patterns. Many classification models can be considered and compared in the analysis. It is evident that law enforcing agencies can take a great advantage of using machine learning algorithms to fight against the crimes and saving humanity. For better results, we need to update data as early as possible by using current trends such as web and Apps.

IV. REFERNCES:

1. Mohammed Boukabous, Mostafa Azizi, "Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers", Indonesian Journal of Electrical Engineering and Computer Science Vol. 25, No. 2, February 2022, pp. 1131~1139 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v25.i2.pp1131-1139
2. Ricardo Francisco ReierForradellas, Sergio Luis Náñez Alonso, Javier Jorge-Vazquez and Marcela Laura Rodriguez, "Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction", Soc. Sci. 2021, 10, 4. https://doi.org/10.3390/socsci10010004
3. Neil Shah, Nandish Bhagat and Manan Shah, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention", Visual Computing for Industry, Biomedicine, and Art (2021) 4:9, https://doi.org/10.1186/s42492-021-00075-z
4. Panagiotis Stalidis, Theodoros Semertzidis and Petros Daras, "Examining Deep Learning Architectures for Crime Classification and Prediction", Forecasting 2021, 3, 741–762. https://doi.org/10.3390/forecast3040046
5. PawełCichosz, "Urban Crime Risk Prediction Using Point of Interest Data", ISPRS Int. J. Geo-Inf. **2020**, 9, 459; doi:10.3390/ijgi9070459
6. Shaobing Wu, Changmei Wang, Haoshun Cao, and Xueming Jia, "Crime Prediction Using Data Mining and Machine Learning", Springer Nature Switzerland AG2020AISC905, pp. 360–375, 2020, https://doi.org/10.1007/978-3-030-14680-1-40
7. Sarah Brayne and Ange`le Christin, "Technologies of Crime Prediction: TheReception of Algorithms in Policing andCriminal Courts", Advance Access Publication Date: 5 March 2020 Article, doi: 10.1093/socpro/spaa004
8. Shraddha Ramdas Bandekar, C. Vijayalakshmi, "Design and analysis of Machine Learning Algorithms for the reduction of crime rates in India", 2020, Elsevier B.V., doi: 10.1016/j.procs.2020.05.018
9. Chao Huang, Junbo Zhang, Yu Zheng, Nitesh V. Chawla, "DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction", 2018 Association for Computing Machinery, doi.org/10.1145/3269206.3271793

10. Shubham Agarwa, Lavish Yadav, and Manish K Thakur, "Crime Prediction based on Statistical Models," Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India.
11. Alkesh Bharati, Dr Sarvanaguru RA.K, "Crime Prediction and Analysis Using Machine Learning",International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 09 | Sep 2018 www.irjet
12. Nelson Baloian1, Enrique Bassaletti, Mario Fernández, Oscar Figueroa, Pablo Fuentes, "Crime Prediction using Patterns and Context", Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design
13. Emmanuel Ahishakiye, Elisha Opiyo Omulo, DanisonTaremwa, Ivan Niyonzima, " Crime Prediction Using Decision Tree (J48) Classification Algorithm", International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 06 – Issue 03, May 2017
14. Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning Algorithms To Analyze Crime Data", Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015
15. Shamsuddin, N. H. M., Ali, N. A., &Alwee, R, " An overview on crime prediction methods", In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
16. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM- Based Cloning and Jamming Attack Detection in IOT Sensor Networks."Advances in Information Communication Technology and Computing, pp. 461-471. Springer, Singapore, 2022 .https://link.springer.com/chapter/10.1007/978-981-19-0619-0_41
17. Thangamani, M., and S. Jafar Ali Ibrahim. "Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso- Wc) and Gene Ontology for Microarray Gene Expression."In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, pp. 48-55. 2018.https://dl.acm.org/doi/abs/10.1145/3299852.3299866

Historical Analysis And Forecasting Of Stock Market Using Fbprophet

N. Suresh[1], B. Priya[2], M. Sai Teja [3], B. Sai Krishna[4], G. Lakshmi [5]

Assistant Professor (IT), Dept. of Information Technology, QIS College of Engineering & Technology, India (1).

Final Year (B.Tech), Dept. of Information Technology, QIS College of Engineering & Technology, India (2,3,4,5)

ABSTRACT:

Forecasting can be used in many fields such as crypto currencyprediction, financial entities, supermarkets etc. We get the time series datewhich we use to feed the data into the algorithm is given by Y finance withthis we get refreshed data every day. The stock market prediction or for ecasting helps customers and brokers get a brief view of how the market be haves for the coming years. Many models are currently in use Like [[Regression techniques, Long Short-Term Memory algorithmetc .FB Prophetis provent perform better than most other Algorithms with better accuracy. From the proposed research and references we have determinedFacebook's Prophet algorithm as our forecasting algorithm because it ispredicting at better accuracy, low error rate, handles messy data, doesn't bother for null values and better fitting.

1. INTRODUCTION:

The technique of predicting the future value of a company's shares of an exchange is known as stock market prediction. Our project aims atforecasting the future of the stock market on a specified day or over theperiod of time. This project will help the stock market enthusiasts to help predict the stock price and hence helping those people in getting profits. In order to do all the, operations we need an algorithm that could help us in doing all the analysis. We have to choose analgorithm that could be flexible to holidays and thus, helping us in analysing seasonal data of a stock. Hence, we

chose FBP roph eta so algorithm as it can be flexible to messy data and can also handle null values very well. The major goal of our project is to predict the future value of the financial stocks of a company. The project aimisto help the users get correct prediction of stock and helps him better understanding of the stock market and leads him to huge profits.

The dataset we use here is from Y finance which provides historicaldata of a stock company since its inception. Data from our dataset gets refreshed every day. The ultimate goal of the proposed project was toForecast the future value of the financial stocks of a company with respect to date. We make use of Machine learning to for ecastina better way. There are many algorithms provided but we felt FB Prophet will dot he job better. It is being used in some of the applications in Facebook. Also, we are using a web frame work for python called stream lit through which we can deplo your project in public medium so that anyone with an URL can view our project which we hoste dinour GitHub.

2. OBJECTIVE:

The major goal of our project is to project the company's share value ofthe future. The project aim is to help the users get a correct prediction of stockand helps him better understanding of the stock market and leads him hugeprofits.

Our project aims at forecasting the future of the stock market on aspecified day or over the period of time. This project will help the stockmarket enthusiasts to help predict the stock price and hence helping thosepeople in getting profits. In order to do all the, operations we need analgorithmthat couldhelpusindoingall the analysis.

3. EXISTING SYSTEM:

There are many existing stock market forecasting applications available but all of those are for either investors or to make money viaadvertising revenues. Some of the disadvantages of existing systems are poor use rinter face, low level of accuracy, too many ads, asking for personal information touse their services, etc.

4. PROPOSED SYSTEM:

- Using Y Finance module, the historical stock market data is collected.
- The obtained data is then trained in the FB Prophet algorithm.
- Using the FB Prophet algorithm, we forecast the stock market of thefuture.
- An interactive and User-friendlyd ash board is then created.
- Then the project is Cloud enabled using Heroku cloud and hosted using GitHub and then using an URL the application can be accessed anywhere.

5. LITERATURE SURVEY:

- **'Time series forecasting model for supermarket sales using fb prophet'** by Shilpapande, Bineetkumarjhain2021 their paper aimsinimple menting he Facebook's Prophetalgorithm, conducting ananalysis of sales related to supermarket with provided data. fitting the obtained data in to the ARIM A model for analysis.

- **'Stock Price Forecasting Using Data from Yahoo Finance and Analyzing Seasonal and Nonseasonal Trend'** by Jai Jagwani, ManavGuptain2018. This paper offers a bigger picture of share value projection by merging the outputs of multiple analysis models in an attempt to implement arrange of share value prices for brokers such that no losses are incurred.

- **'Applications of Facebook's prophet algorithm for successful sales forecasting based onreal-world data'** by Kemal Korjenic, Emirzunic in 2020. This journal provides a system for reliably projecting future retail sales and classifying product offering according to predicting accuracy standards in there tail industry.

- **'Time Series Model for Stock Market Prediction Utilizing Prophet'** by Neha Gupta,Dr.LillySheebaSin2021.This journal tries to highlight the importance of timing inenhancing fore casting accuracy, which is accomplished with the use of prophet algorithm. This journal uses fb prophet library to define 3 different hyper parameters namely seasonality, trend, and holidays.

- **'Stock price forecasting using information from Yahoo Financeand google trend'** by Selene Yue Xu in 2014. This publication aims toanticipate periodic fluctuations in share price using a combination oftraditionalseries dataanalysis anddata outfromYahoo.

6. IMPLEMENTATION:

DATA COLLECTION:

To develop a better framework for our model to forecast and predict stock values we have explored many kinds of data set sin google and Kaggle. Eventually we came to a conclusion that YFINANCE will give better datawhich gets refreshed every day. The data is well structured and containsevery detail about a company. Our dataset contains lots of columns which are very commonly used terms for stockmarket analysts and enthusiasts.
Date-Historic date [of every stock traded day

Open- The starting period of trading on a securities exchange or organized over-the-counter market.

High- The high est price at which a stock traded during the course of the trading day.

Low- Minimum price of a stock in a period.

Close-The price of the last traded stock attheend of the day.

Volume - The number of shares traded in a stock or contracts traded infuturesor options.

Dividends: The distribution of some of a company's earnings to a class of its share holders, as determined by the company's board of directors.

Stock splits- Astock split is when a company's board of director is suesmore shares of stock to its current shareholders without diluting thevalueof their stakes.

6.2. ARCHITECTURE:

Our project could project stock market data of the future more accurately when compared to all other [ecasting algorithms such as ANN,ARIMA etc. [This would be really helpful for those who are willing to invest instocks and for the brokers too. Yfinance is a module in python, whichprovides daily market data using an API from Yahoo. We propose a basic flow of method ology with a [small block diagrams hown [below.

6.2.1. Creating Our Forecasting Model:

There are many forecasting models available in the market like LSTM, Regression models, ANN etc., But then ewly introduced FB Prophet shows better results compared to all the other algorithms. FBProphet is a powerful tool being used by Facebook even in its manyforecasting applications. It has also been noted that this algorithm in mostof the cases. Also, some of the main reasons for applying this model arethis algorithm can handle messy data, can work if data contains lot of nullvalues also provides special feature to exclude holidays. The algorithm makes use of Pystan which is as tate of art tool for statistical analysis using probabilistic algorithms.

6.2.2. User Interface:

Now, we need to implement a User Interfacet deployour application into web. Stream lit is an opensource framework in python that provides users to create and share their applications in the web.It is one the most used

frameworks for data science. Using this frame work we caneasily deploy our application in the public domain and host it in any cloud mediumsor GitHub.

6.2.3. Creating Interactive plots:

We try to implement the data we get from Y finance in several plots and graphs.The users prefer data representation in graphs and plots rather than verbal representation. So, we try to develop an interactive visualization dashboard designed to show the seasonal trends and forecasting of thestocks of the future. We are making use of several packages such as Ploltyand some methods that are provided by the FB Prophet. Python allows dynamic and interactive graphs. Using those graphs, we could get extra information while we hover over the graph and can easily view detailedbehaviour by zooming in and out. These interactive plots can be used to represent various important features of a stock industry such as open, close, high, low, dividends etc. We implement dates on the x-axis and values on they-axis.

6.2.4. Fore casting and Seasonal Data Analysis:

Using FB prophet we try to predict the data for the coming years. We trainthe FB Prophet model using the Historical data we get from Y finance andthen predict the data of the future dates. This algorithm provides resultswith better accuracy. Also, the algorithm works best with time series data and provides various techniques like excluding holidays and weekly,monthly, daily, yearly analysis. the algorithm also provides methods to display plots for those seasonal analysis.

6.2.5. Project Deployment:

After the model has been completed successfully and user interface hasbeen created then we will our application cloud enabled and host using GitHub. Finally, we deploy our project instream lit after that stream lit provides a link with which any one can view our project. Also, our application can be shared to anyone using emails so that they can use our applications. Currently, our application is in public domain we can also make our application for private Use.

7. CONCLUSION:

We believe our project clearly ai msat projecting the share value of a company of the future. Therefore, we are aiming to implement a web-based data science application using stream lit, frame work, and we try to enable our application in cloud we will host our application in git hub, so that anyone for many where can access our application using an URL.

8. REFERENCES:

1. Dzenana, Emir & Korjenić,Kerim&Donko, Kemal & Hodžićand Zunic "Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Basedon Real-world Data, "International Journal of Computer Science and Information Technology (IJCSIT), Vol 12, No 2, pp. 23-36,April 2020
2. ChenGuo,Quan bo Ge, HaoyuJiang, Gang Yaoand QiangHua, Maximum "Power Demand Prediction Using Fb prophet with Adaptive Kalman Filtering", IEEE Access, 2020.
3. Ahmed Tealab, "Time series for ecasting using artificial neural networks methodologies: A systematic review", Future Computing and Informatics Journal, vol.3,Issue2, pp.334-340,December2018.
4. BineetKumarJha,Shilpa Pande "Time Series Forecasting Model for Super market Salesusing FB-Prophet" IEEE2021.
5. Dr. Lilly Sheeba. S, Neha Gupta "Time Series Model for StockMarket Prediction Utilising Prophet"TJCME2021.
6. Emir Zunic, Kemal Korjenic "Application Of Facebook's Prophet Algorithm For Successful Sales Forecasting Based On Real-World Data"IJCSIT 2020.
7. JaiJagwani,Manav Gupta,Hardik Sachdeva"Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend"IEEE2018.
8. Selene Yue Xu "Stock Price Forecasting Using Information fromYahoo Finance and Google Trend",UC Brekley,2014.

9. Ilaria Bordino, Nicolas Kourtellis, Nikolay Laptevand Youssef Billawala"StockTrade Volume Prediction wirh Yahoo Financeuser Browsing Behaviour",IEEE2014.
10. Suhas Patil, Rajesh Argiddi and SulabhaApte "Financial forecasting by improved fragmentation algorithm of Granular Fragment basedmining",IEEE2015.
11. Ferdiansyah Ferdiansyah, Siti Hajar Othman, Raja Zahilah Raja MdRadzi,DerisStiawan,Yoppy Sazakiand Usman Ependi"ALSTM-Methodfor Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market",IEEE2019.
12. Jafar Ali Ibrahim. S, Mohamed Affir. A *"Effective* Scheduling of Jobs Using Reallocation of Resources Along With Best Fit Strategy and Priority", International Journal of Science Engineering and Advanced Technology(IJSEAT) – ISSN No: 2321- 6905, Vol.2, Issue.2, Feb-2014, http://www.ijseat.com/index.php/ijseat/article/view/62
13. C. Narmatha , Dr. M. Thangamani , S. Jafar Ali Ibrahim, " Research Scenario of Medical Data Mining Using Fuzzy and Graph theory", *International Journal of Advanced Trends in Computer Science and Engineering, Vol 9, No 1 (2020): 349-355,* https://doi.org/10.30534/ijatcse/2020/52912020

Bitcoin Price Prediction Based on Linear Regression and LSTM

Dr. R. Suneetha Rani[1], D. Sri Vinithri Chowdary[2], Ch. Uday Kiran[3], K. Deekshith[4], V. Alekhya[5]

[1]Professor & Head of the Department, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

[2,3,4,5] Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering andTechnology, Andhra Pradesh, India

ABSTRACT

Machine learning based on Neural Network has integrated usages in a variety of fields such as translation, finance, distribution, and medical world as well as cognition. This explains the working of the linear regression and Long Short-Term Memory model in predicting the value of a Bitcoin. Due to its raising popularity, Bitcoin has become like an investment and works on the Block chain technology which also gave raise to other crypto currency. This makes it very difficult to predict its value and hence with the help of Machine Learning Algorithm and Artificial Neural Network Model this predictor is tested. This study shows Artificial Neural Network Learning Model on the basis of LSTM, which analyses the previous prices of a cryptocurrency, Bitcoin and predicts the next one. Bitcoin is a crypto currency used worldwide for digital payment or simply for investment purposes.

Keywords- Bitcoin, Block chain, Crypto currency, Linear Regression, LSTM

1. INTRODUCTION

Bitcoin is a digital crypto currency that operates on an online decentralized network; it can be traded using an online peer-to-peer Bitcoin network that is not reliant on a central bank or a single administrator. Because it is accepted in over 40 countries worldwide (including Germany, Canada, and Croatia), the emergence of new alternative coins has resulted from its growing popularity. Bitcoin is also used to exchange other crypto currencies, products, and services. Since the introduction of this crypto currency in the year 2009, no hacker has been able to infiltrate it due to block chain technology, where each

electronic coin is encrypted with a unique digital signature which makes it easier to track and can be trusted. Each owner signs a digital hash from the previous transaction and adding the public key of the next owner before passing it. The price of Bitcoin in January 2017 was 1,000USD and by the end of December 2017, its value went up to 16000 USD and its value as on July 2021 is 32818USD. We can say that the crypto market is very volatile, and among all the crypto currencies in the market, Bitcoin is experienced by most of the investors due to its anonymity and transparency in the system. This research aims to work on the prediction system for Bitcoin using various Machine learning algorithms and deep learning models to predict the price. There are various factors affecting the price of Bitcoin, in this project we will focus on open, close, high, and low factors.

This paper also consists of 6 libraries:

- **Pandas**: Provides an environment to python for creating creative and practical statistical computing for financial data analysis applications.
- **Seaborn**: Effective visualization for better understanding the graphs and charts.
- **Scikit-Learn**: Implementation of various algorithms. It contains a large variety of supervise and unsupervised learning algorithms.
- **Tkinter**: To create a faster and quicker GUI application, with cooler features including the implementation of CSS support.
- **Pickle**: Serialization and de-serialization of python object structure to store it in a file/database, maintaining program state, and transfer of data over a network.

2. RELATED WORK

The prediction of crypto coins using the SVM and SVM-PSO method is suggested, where they used the day trading method to predict the values of ETH, BTC, XEM, XRP, XLM, LTC. SVM-PSO shows the optimized results. Performance accuracy of different Classifiers differs from coin to coin. However, this paper works only with a machine learning algorithm, and hence

the data can be furtherimproved by implementing the Deep Learning concept. The prediction of Bitcoin price using a transaction graph is proposed. The experiment consists of the Baseline, Logistic Regression, SVM, and Neural Network model with an accuracy of 53.4%,54.3%, 53.7 and 55.1%. The feature selection in this paper is based on the Bitcoin block chain network which tends to be the least informative feature for the prediction of the Bitcoin price. Predicting the crypto currency prices using sentimental analysis and Machine learning concepts like SVM and Random Forest on ETH, BTC, and XRP with BTC being the highest accuracy of 0.72. This accuracy rateis very low since machine learning algorithms were applied and it can be improved by testing with deep learning models. A prediction model was proposed using four major algorithms, Gradient Boosted Tree, Neural Network, Ensemble Learning Method (with the best accuracy of 92.4%). The prediction system using Log regression, SVM, ANN, and random forestwas proposed and shows that SVM has thebest accuracy regarding a time-scale activity consisting of daily, 15-, 30- and 60-minutes return. Although SVM does tend toshow better results out of the 4 algorithms, the prediction system can still show better results when Deep learning concepts are applied. A linear regression model was usedto predict the various cryptocurrency price using the open, low, and high cost. The experiment shows an accuracy of 99.3%. This paper does consist of a high accuracy rate but the data set used is comparatively small for a model to work on a real-time chart.

3. OBJECTIVE

The objective of this proposed system is to develop an application which will predict the bitcoin prices in future with decent accuracy. This allows the investors to invest wisely in bitcoin trading as the prices of bitcoin have gone up to an exaggerating amount in the last ten years.

Thus, the main objectives of the "Bitcoin Price Prediction" can be stated as follows:

1. Develop an application which will predict the bitcoin prices in future with decent accuracy.
2. Allow the investors to invest wisely in bitcoin trading as the prices of bitcoin havegone up to an exaggerating amount in the last ten years.
3. Make use of machine learning algorithms to increase the accuracy of Bitcoin price prediction.

4. LIMITATION

Although crypto trading has become a new trend, the increase in the number of digital coins and the adaptation of block chaintechnology causes the biggest concern i.e., scalability. It is still dwarfed by the number of transactions that, VISA, processes each day. Additional to that is the speed of transaction which the crypto market cannot compete with the players like VISA and MasterCard until the infrastructure delivering these technologies is massively scaled. The crypto market is very volatile and can never be predicted at 100 percent accuracy. The market depends on human sentiment too; you may never know when a person owning at least 100 Bitcoin can suddenly sell his entire asset and create a big dip in the crypto market. We can never predict a human emotion even with the advanced technology we have in hand. The analysis of any technical chart composes of mainly 3 major topics, the trend and momentum which indicate the direction and strength of direction, support, and resistance which indicates the potential stopping points of those directions, and the pattern in general, which indicates the information about the market psychology. Cryptocurrencies have not been around forlong enough to provide sufficientinformation regarding the resistance andkey support compared to the stock market, currencies, and commodities. This makes it difficult to predict and practice.

5. EXISTING SYSTEM

A Cryptologic pioneer, David Chaum, devised Blind Signature technology, which telecommunicates the encoded messages sealing digital signature, and resulted in inventing Ecash. That is the primary commercial cryptocurrency. Bit 29Coin in 2009 was the new cryptocurrency accomplished with Block Chain technology. After that most cryptocurrencies have been improved with Block Chain based. Ethereum emerged as the developed money which has services and applications in addition to Block Chainsystem in 2015. WEF (World Economic Forum) suggested that the ranking of Blockchain must be the fourth of 12 future technologies in the Global Risks Report. Furthermore, in 10 years, 10 percent ofGDP all over the world is expected to be based on Blockchain technology. In April 2019, about 40 major banks around theworld announced that they would experiment CBDC (central bank digital currencies) founded on Blockchain.Blockchain, which is encrypted with trade information on the public or private network, is a diversified ledger shared withrelevant network participants.

Disadvantages:

1. In existing system Block Chain Technology is used to predict bitcoin price.
2. By using Blockchain Technology the prices may not be constant they may vary day to day.
3. By using Blockchain Technology we cannot predict the future prices.

6. PROPOSED SYSTEM

The proposed model is used to predict bitcoin price using Machine learning and Neural Network. Machine learning uses Linear Regression and Neural Network uses LSTM for predicting bitcoin prices. In Data Segregation we use features like Open, Close, High, low, Volume BST, Volume UDST,

Time, Symbol Linear Regression accuracy rate is 99.87% whereas, LSTM accuracy rate is 97.56%. It is discovered that the Linear Regression model accuracy rate is very high when compared to other models. In this study, we have used data sets for Bitcoin for testing and training the ML and AI model. With the help of python libraries, the data filtration process was done. Python has provided with a best feature for data analysis and visualization. After the understanding of the data, we trim the data and use the features or attributes best suited for the model. Implementation of the model is done and the result is recorded. It was discovered that the linear regression model's accuracy rate is very high when compared to other Machine Learning models from related works; it was found to be 99.87 percent accurate. The LSTM model, on the other hand, shows a mini error rate of 0.08 percent. This, in turn,demonstrates that the neural network model is more optimized than the machine learning model.

7. METHODOLOGY

7.1. Data collection: Data Collection is the first step we take in order to start any project. It is defined as the procedure of collecting, measuring, and analyzing accurate insights for research using standard validation techniques. An analyst would then be able to assess their theory dependent on gathered information. By and large, information assortment is the essential and most significant advance for research, independent of the field of examination. The methodology of information assortment is diverse for various fields of study, contingent upon the necessary data. The most important objective of data collection is ensuring that the gathered information is rich in content and reliable for statistical analysis so that data-driven decisions can be made efficiently and effectively. The data set contains day transactions from 29th August 2017 to 9th August 2020. The data is first tested out with certain regression techniques and then a deep learning model is implemented to provide better accuracy compared to machine learning concepts when there is high or more data sets.

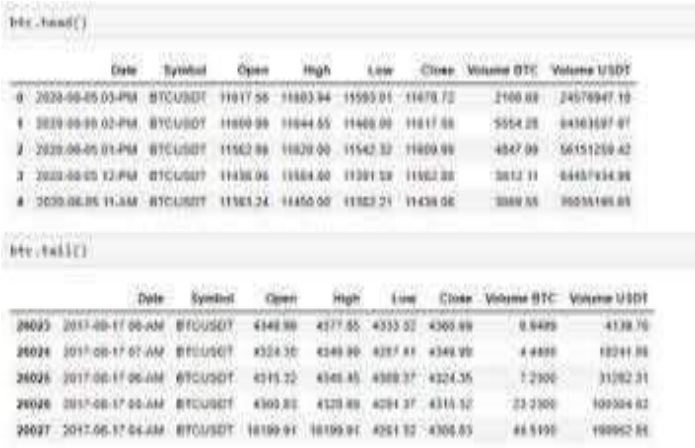


Fig 1. Display of the Data collected

7.2. Feature Selection: Now that we have the required data for the project, we need tostart the next procedure called data segregation or feature selection. This is a process where we trim out the unwanted data or we remove the unnecessary data from the data set. This step is necessary as we require only those features which can contribute to our prediction as unnecessary data can cause noise in our final output. Toput it in simple words, we segregate data so that we can have a better model which provides us with an optimized result, reduce the property of over-fitting or redundancy and reduce the training time so that thesystem can generate output faster and with higher accuracy. In this project, I have implemented a few predefined python libraries which help in data visualization and can help you understand the important features which are required by the system. Data visualization is a technique where data or information is represented in a diagrammatic format for better understanding. Data visualization helps us to communicate with the relationships of data using the help of images. These imagesare in form of patterns that can be understood very easily. This is one of the main reasons how machine learning helps in analyzing data. Whether you work in the finance department or marketing or technical or design, you need to visualize data to understand it. This makes data visualization an important factor in today'sworld.

Features	Definition
Open	Opening value of trade at that time stamp
Close	Closing value of trade at that time stamp
High	Highest trade value in the time stamp
Low	Lowest trade value in the time stamp
Volume BTC	Total trade volume in BTC in the given timestamp
Volume USDT	Total trade volume in USDT in the given timestamp
Date	The given date and time of each bid
Symbol	Symbolic representation of coin

Fig 2. The Features represented in the data

With the help of data visualization libraries, we can see the correlation between features and pinpoint the ones which we require. A sample image is shown below to show the correlation graph between the features in the given data set. You can notice in the given image Figure 3

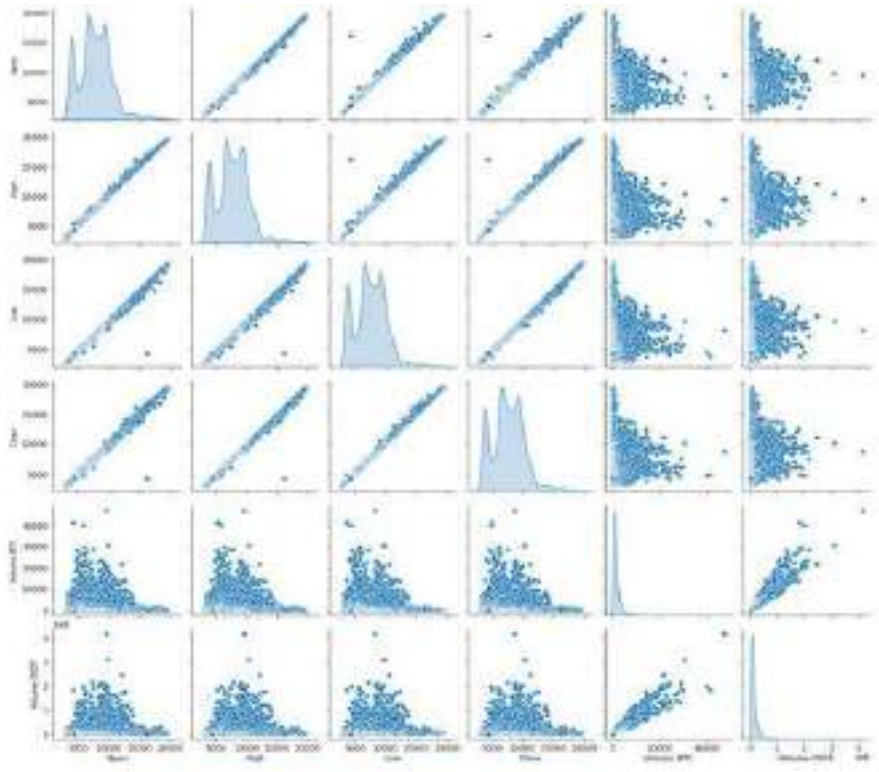


Fig 3. Correlation graph between the features.

7.3. Data Preparation: When variablesthat are measured in different scales it does not contribute equally in model fitting which will lead to model learned function to create a bias. Thus, standardization or normalization of data is very much essential for better accuracy and result. When working with a Machine Learning model or Deep Learning models where we require back propagation to be more stable andeven faster, proper scaling of data is necessary.

$$x_{scaled} = \frac{x - min\ (x)}{max(x) - min\ (x)}$$

8. ALGORITHMS IMPLEMENTED

8.1. LINEAR REGRESSION

This technique is used to identify the relationship between dependent and independent variables and is leveraged to predict future outcomes. When we use only one dependent and one independent variable then it is called the simple linear regression. As the number of independent and dependent variable increase, it is then referred to as multi-linear regression. The graph is plotted using a straight line across the graph which seeks to be the best fit by calculating the method of least square.

y = mx + C

C = y intercept

m = slope

x, y are the points on the graph

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared errorn = number of data points

Y_i = predicted values

$\hat{Y}_i$ = predicted values

8.2. LONG SHORT-TERM MEMORY(LSTM)

It is a deep learning concept or particularly a Recurrent Neural Network concept that avoids the vanishing gradient problem. The main reason for using this algorithm is that it avoids the back propagation error from vanishing or exploding, instead, these errors can flow backward through an unlimited number of virtual layers unfolded in space. LSTM mainly works on time series graphs with data sets that consist of events that occur thousands or millions of discrete-time steps earlier. It works with given long delays between significant events and can also handle signals with a mixture of low and high-frequency components. Over a lot of researchers have used LSTM to predict time series related data sets for stock prediction and have achieved greater or higher accuracy compared to other algorithms. LSTM is capable of recognizing context-sensitive language unlike any other previous models based on Hidden Markov Models (HMM) and other similar concepts.

The main formulation of the result in LSTM is based on Mean Absolute Error, the equation

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

y = prediction

x = true value

n = total number of data points

9. RESULT AND DISCUSSION

After the data analysis process, we find that the only four features were well suited for the testing of this project. The data was trimmed and only the selectedfeatures were left as shown in Figure 4.

	Open	High	Low	Close
0	11617.56	11693.94	11593.01	11678.72
1	11609.99	11644.65	11466.00	11617.56
2	11562.86	11620.00	11542.32	11609.99
3	11438.06	11584.60	11391.59	11562.86
4	11393.24	11450.00	11382.21	11438.06

Fig 4. Attribute/Features selected are Open,High, Low, and Close

We can see the output of two models, one which is the Machine Learning model i.e., Linear regression, and the other one is the Recurrent Neural Network model i.e., Long Short-Term Model which shows us the two different outcomes. Linear regression tends to work based on the Mean Squared Equation which tells us the accuracy of the linear graph with respect to the continuous-time frame data set. We see that the accuracy of the training data is approximately 99.97% and the accuracy of the testing data is tending to be approximately 99.97% as shown in Figure5. Meanwhile, the LSTM model tends to find the accuracy with respect to the Mean Absolute Error which shows the error rate approximately to be 0.08% as shown in Figure6.

```
In [38]: model.score(x_train, y_train)
Out[38]: 0.9997158887216999

In [39]: pred = model.predict(x_test)

In [40]: model.score(x_test, y_test)
Out[40]: 0.9997966090479169
```

Fig 5. Accuracy obtains from the training and testing data set using Linear regression model

S.No	Open	High	Low	Close	Expected Result
1	11617.56	11693.94	11593.01	11678.72	11669.05
2	11609.99	11644.65	11466.00	11617.56	11530.74
3	11562.86	11620.00	11542.32	11609.99	11603.13
4	11438.06	11584.60	11391.59	11562.86	11525.64
5	11393.24	11450.00	11382.21	11438.06	1140.47

Fig 6. Testing of Linear regression model

Fig 7. Final Resultant graph of LSTM and theMean Absolute Error rate (0.08%)

DISCUSSION:

The Data visualization shows the correlation between all the features and only the four selected features have a sharp correlation. Data is then fitted into the model using the predefined commands accessible to python. These data modelswere trained and tested out with a limited number of data sets and provided the result.

With the growing technology and the raise in the data sets we can still work on the model with various other alternative crypto currencies. The model shows a better prediction rate for LSTM but with a very slight difference compared to the linear regression model.

10. CONCLUSION

According to the findings, Long Short-Term Memory has a higher accuracy rate than Linear Regression. Because this study only compares the

features of open, close, high, and low, the outcome may alter if we examine additional factors. Data sets cannot be the main rationale for forecasting because the crypto market is dynamic and influenced by social media and other external factors. New data can be acquired, evaluated, and rehearsed as technology progresses, resulting in greater findings forthis experiment.

FUTURE SCOPE

1) More algorithms are being implemented in order to determine the best approach for predicting the crypto currency.
2) Implementing IOT model for smart automatic analysis.
3) To work on a better User Interface so that people can access these data easily and effortlessly.

11. REFERENCES

1. Sin E, Wang L. Bitcoin price prediction using ensembles of neural networks. In: 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery. IEEE. 2017;p. 666–671. doi:10.1109/FSKD.2017.8393351.
2. Shankhdhar A, Singh AK,Naugraiya S, Saini PK. Bitcoin Price Alertand Prediction System using various Models. *IOP Conference Series: Materials Science and Engineering*. 2021;1131(1):012009.Available from: https://dx.doi.org/10.1088/1757-899x/1131/1/012009.
3. Mittal R, Arora S, Bhatia MP. Automated cryptocurrencies prices prediction using machine learning. *ICTACT Journal on Soft Computing*. 2018;8(4):1758– 1761. Available from: http:// ictactjournals.in /paper/IJSC_Vol_8_ Iss_4_Paper_8_1758_1761.pdf.
4. Nakamoto S. Bitcoin: A peer-to- peer electronic cash system. 2019. Available from: https://git.dhimmel.com/ bitcoin -whitepaper/.
5. Sebastião H, Godinho P. Forecasting and trading cryptocurrencieswith

machine learning under changing market conditions. *Financial Innovation*. 2021;7(1):1–30. Available from:https://dx.doi.org/ 10.1186/ s40854-020-00217-x.

6. Phaladisailoed T, Numnonda T. Machine learning models comparison for bitcoin price prediction. *10th InternationalConference on Information Technology and Electrical Engineering*. 2018;p. 506–511. Available from: 10.1109/ICITEED.2018.8534911.
7. Jaquart P, Dann D, Weinhardt C. Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*. 2021;7:45–66. Available from:10.1016/j.jfds.2021.03.001.
8. Rane PV, Dhage SN. Systematic erudition of bitcoin price prediction usingmachine learning techniques. *5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. 2019;p. 594–598. Available from: 10.1109/ICACCS.2019.8728424.
9. Roy R, Roy S, Hossain MN, Alam MZ, Nazmul N. Study on nonlinear partial differential equation by implementing MSEmethod. *Global Scientific Journals*. 2020;8(1):1651–1665.
10. Mckinney W. Pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*. 2011;14(9):1–9.Available from:https://www.dlr.de/sc/portaldata/15/resour ces/ dokumente/ pyhpc2011/submissions/pyhpc2011_submission_9.pdf.
11. Waskom M. Seaborn: statistical data visualization. *Journal of Open Source Software*. 2021;6(60):3021. Available from: https://dx.doi.org/10.21105/joss.03021.
12. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuro informatics*. 2014;8:14. Available from:https://dx.doi.org/10.3389/fninf.2014.00014.

13. Polo G. PyGTK, PyQT, Tkinter and wx Python comparison. *The Python Papers*. 2008; 3:26–37. Available from: https:// www. mclibre.org /descargar/docs/ revistas/the-python-papers/the-python- papers-3-1-en-200804.pdf.

Insulin Dosage Prediction System For Diabetic Patients

Neelam Sai Kiran [1], M. Keerthana [2], P. Bhavani [3], D. Revanth [4], P. Siva prasad [5]

Assistant Professor (IT), Dept. of Information Technology, Qis College of Engineering and Technology, Andhra Pradesh, India (1)

Final Year (B. Tech), Dept. of Information Technology, Qis College of Engineering and Technology, Andhra Pradesh, India (2,3,4,5)

ABSTRACT: Diabetes is a common chronic disease and poses a great threat to human health. Diabetes can lead to chronic damage and dysfunction of various tissues, especially eyes, kidneys, heart, blood vessels and nerves. To control the BGLs of diabetic patients the traditional prevention techniques such as eating healthy food and conducting physicalexercise plays a major role and also taking the proper amount of insulin dosage has the crucial rule in the treatment process.The rapid development of machine learning has been applied to many aspects of medical health.We have proposed a model based on artificial neural network (ANN) to predict the proper amount of insulin needed for the diabetic patient.The proposed model was trained and tested using several patient's data containing many factors such as weight, fast blood sugar and gender.

1. INTRODUCTION:

Diabetes mellitus is a chronic disease characterized by hyperglycemia. Normally, by adjusting of blood glucose levels (BGLs), diabetic patients could live a normal life without the risk of having serious complications. However, blood glucose levels of most diabetic patients are not well controlled for many reasons. The characteristic of diabetes is that the blood glucose is higher than the normal level, which is caused by defective insulin secretion or its impaired biological effects, or both. According to the growing morbidity in recent years, in 2040, the world's diabetic patients will reach 642 million, which means that one of the ten adults in the future is suffering from diabetes.Diabetes can be

divided into two categories, type 1 diabetes (T1D) and type 2 diabetes (T2D). Patients with type 1 diabetes are normally younger, mostly less than 30 years old.Type 1 diabetes is an autoimmune disease in which the beta-cells of the body are destroyed, thus resulting in a lack of insulin production. The typical clinical symptoms are increased thirst and frequent urination, high blood glucose levels. This type of diabetes cannot be cured effectively with oral medications alone. Due to the lack of insulin production, type 1 diabetes patients are required to take insulin subcutaneously as their primary method of therapy. Type 2 diabetes occurs more commonly in the middle-aged and elderly people, which is often associated with the occurrence of obesity, hypertension, dyslipidemia, arteriosclerosis, and other diseases.

One of the essential components of a diabetes management system concerns the predictive modelling of the glucose metabolism. It is evident that the prediction of glucose concentrations could facilitate the appropriate patient reaction in crucial situations such as hypoglycaemia.

Thus, several recent studies have considered advanced datadriven techniques for developing accurate predictive models of glucose metabolism.

2. OBJECTIVE:

The aim of this project is to predict insulin dosage for diabetic patients. The main objective of this project is to predict whether the person has Diabetes or not based on various features like height, weight, fast blood sugar, gender. If the person is a diabetic patient, then using the Back Propagation(BP) Algorithm in Artificial Neural Network (ANN) model, we can predict the proper amount of insulin dosage required for that particular patient.

In addition to the general guidelines that the patient follows during his daily life, several diabetes management systems have been proposed to further assist the patient in the self-management of the disease. One of the essential components of a diabetes management system concerns the predictive modeling of the glucose metabolism.

3. EXISTING SYSTEM:

ANNs are non-linear mapping structures that are inspired by the function of the human brain and are considered

Power ful modeling tools especially for data with unknown underlying relationships. ANNs consist of computational elements called neurons, operating in parallel and connected by links with variable weights which are typically adapted during the learning process (Fig.1). In our model, ANNs are non-linear mapping structures that are inspired by the function of the human brain and are considered powerful modeling tools especially for data with unknown underlying relationships. ANNs consist of computational elements called neurons, operating in parallel and connected by links with variable weights which are typically adapted during the learning process (Fig.1). In our model,

In existing system KNN (K- Nearest Neighbors) Machine Learning model is used to predict the insulin dosage for diabetic patients.KNN algorithm can be used for both classification and regression problems. The KNN algorithm uses 'feature similarity' to predict the values of any new data points.A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbours.Another approach uses an inverse distance weighted average of the K nearest neighbours. KNN regression uses the same distance functions as KNN classification.

The main disadvantage of this existing system is KNN is not efficient for large datasets, so that we go for ANN in proposed model.

4. PROPOSED SYSTEM:

In the proposed system we are using Artificial Neural Network with Back Propagation algorithm to classify diabetic patients and predict the proper amount of insulin dosage required for the diabetic patient. To train and test the model we used a collection of Electronic Health Record (EHR) data.The input data for our model are: length for patient (cm), weight for patient (kg), fast blood sugar reading for patient (mmol/l) and gender of patient (female/male). So, we can buildan ANN based model that is trained using backpropagation algorithm.

Length (cm)	Weight (kg)	Blood sugar (mmol/l)	Gender (f/m)

5. LITERATURE SURVEY :

Robertson et al. demonstrated Elman's recurrent artificial neural network (ANN) based on meal and insulin intake. The data set originated from a free, artificial mathematical diabetes simulator called AIDA that modelled 28 days of measurements of a T1D patient. Regarding the meal intake, only carbohydrate quantities were considered, and the results are based on the quite limited food absorption modelling capabilities of AIDA.

Another, neural network-based solution is presented by Shanthi and Kumar [13]. The difference between their work and the previously mentioned ANN-based tests is that in this case, the validation data history included real patients in a hospital setting with different insulin therapies using Medtronic's CGMS.

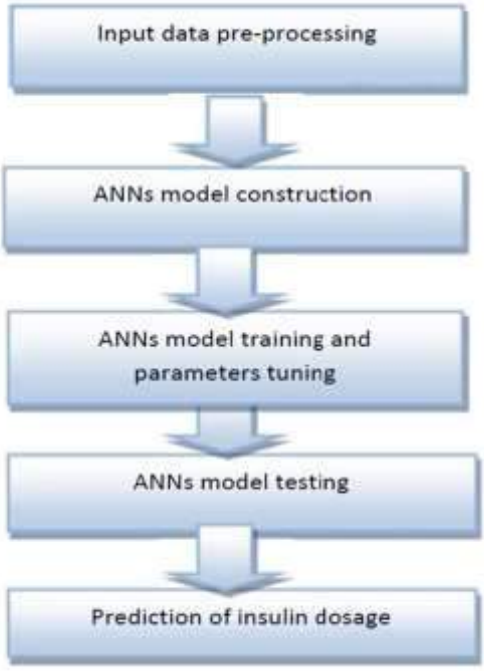
Deeraj Shetty et al. proposed diabetes disease prediction using data mining assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, they propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbour) to apply on diabetes patient's database and analyse them by taking various attributes of diabetes for prediction of diabetes disease.

6. IMPLEMENTATION:

6.1 MODULES:

- Data Collection (EHR data)
- Data Processing
- Model Construction(ANN - Back propagation)
- Training and Testing the Model
- Model Evaluation (Insulin dosage Prediction)

6.2. SYSTEM DESIGN:



6.3 METHODOLOGY:

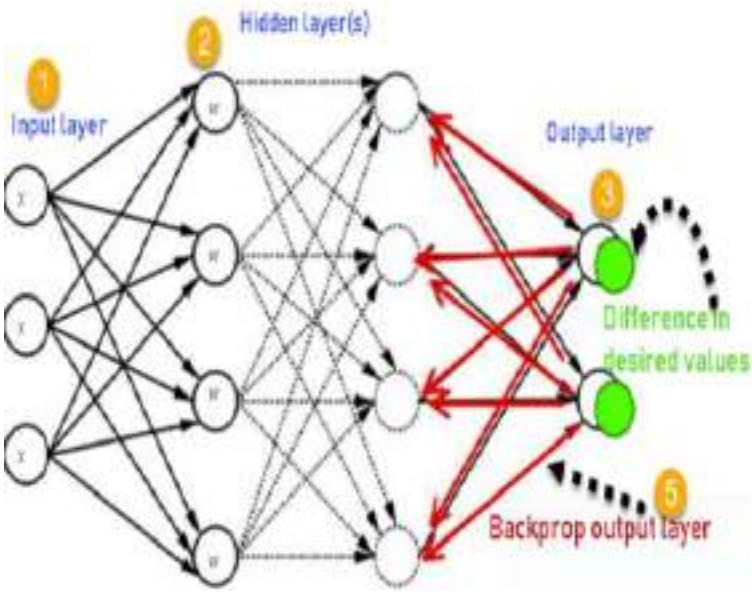
ARTIFICIAL NEURAL NETWORK:

A neural network is a group of connected I/O units where each connection has a weight associated with its computer programs. It helps you to build predictive models from large databases. This model builds upon the human nervous system. It helps you to conduct image understanding, human learning, computer speech, etc.

BACK PROPAGATION:

Back propagation is the essence of neural network training. In 1961, the basics concept ofcontinuous backpropagation was derived in the context of control theory by J. Kelly, Henry Arthur, and E. Bryson. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous iteration. Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

6.4. ARCHITECTURE:



The architecture of back propagation resembles a multi-layered Feed Forward network.

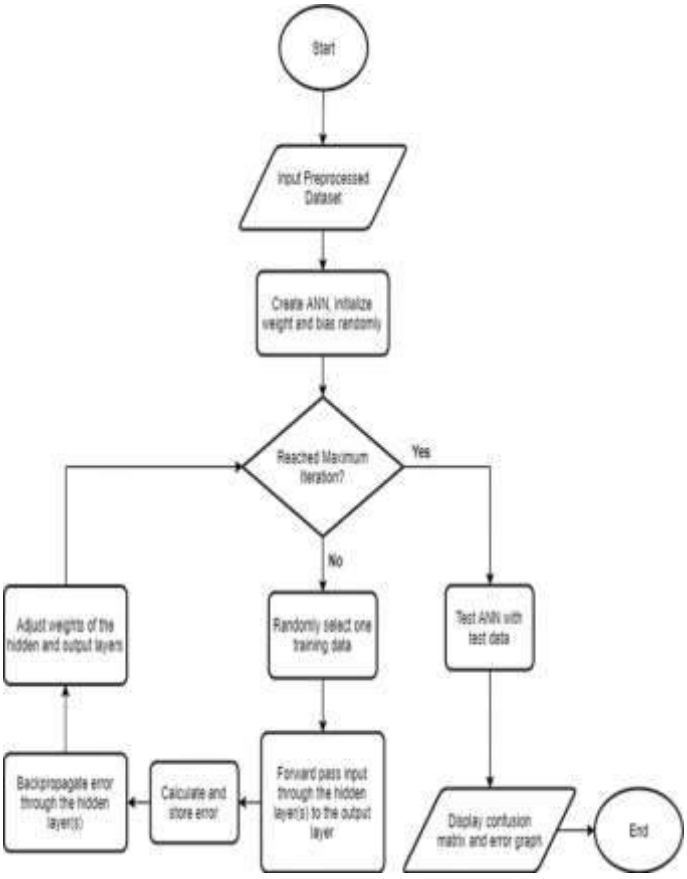
Input Layer: Accept the data (features) and pass it to the rest of the network.Hidden Layer: It is responsible for the excellent performance and complexity of neural networks. They perform multiple functions at the same time such as data transformation, automatic feature creation, etc.

Output Layer:The output layer holds the result or the output of the problem. This method helps calculate the gradient of a loss function with respect to all the weights in the network.The Back propagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The weights that minimize the error function is then considered to be a solution to the learning problem.

The Training algorithm of Back Propagation involves 4 stages:

1. Initialization of weights
2. Feed Forward
3. Back Propagation of errors
4. Updating of the weights and biases

6.5. ALGORITHM:



6.6. WORKING:

We collected our data for 180 patients. These data include: length for patient (cm), weight for patient (kg), fast blood sugar reading for patient (mmol/l), gender of patient (female/male) and the insulin dosage for that patient. The data was divided into two parts: the first part includes 120 reading and was used for training the ANN; the other part includes 60 reading and was used for testing the proposed model.

Step 1: Input data pre-processing: The input data for our model are: length for patient (cm), weight for patient (kg), fast blood sugar reading for patient (mmol/l) and gender of patient (female/male). The output is the suitable insulin dosage for the patient. All input data are normalized in the range (0.0 to 1.0).

Step 2: ANNS model construction: Two thirds of the data are selected to train the model and the other third is used to test it. The proposed prediction algorithm is constructed, it consists of 3 layers: an input layer, a single hidden layer and an output layer. The hidden and output layers use sigmoid activation function.

Step3: ANNS model training and parameter tuning:
In the proposed 3-layer neural network, the number of nodes in the input layer is set to 4, the number of nodes in the hidden layer is varied from 5 to 10 and the learning rate is varied from 0.1 to 0.9. The number of neurons of the hidden layer is set to 7, the learning rate is set to 0.5, and the number of neurons in the output layer is set to 1, as a result, this proposed ANNs model achieves the best performance. The best ANNs model with the suitable number of nodes is selected accordance to the minimum prediction error.

Step 4: ANNs model testing: One third of the data are used to test the accuracy of the proposed prediction model.

Step 5: Prediction of insulin dosage: After training and tuning the proposed prediction algorithm, it can be used to predict new unknown insulin dosage suitable for the patients. MATLAB R2011software was used for the implementation of the proposed model.

Two performance measures related to the prediction errors (PE) were computed.

PE is calculated using the following error equation:

$$PE = |Arv - Prv|\ Arv \quad (1)$$

Where PE is the prediction error, Arv is the actual insulin dosage value, Prv is the predicted insulin dosage value, and || is the absolute value.

Moreover, the prediction accuracy is defined as follows:

$$PA = (1 - PE)\ x100\% \quad (2)$$

Where PA is the prediction accuracy.

CONCLUSION:

This project was aimed at modeling neural network for the prediction of amount of insulin dosage suitable for diabetic patients. A model based on ANN trained with Back Propagation Algorithm was used. The model uses four input information about each patient its length, weight, blood sugar, and gender. We are using Artificial Neural Networks to predict diabetes and applying Back Propagation Algorithm to predict proper amount of insulin dosage required for a diabetic patient.

FUTURE ENHANCEMENT:

The results of the research conducted in this thesis are promising for a wide range of applications in T1D therapy. Nevertheless, the performance and safety of the predictions can be improved further by generating a set of interchangeable models that predict useful Blood Glucose values for control and therapy purposes based on the determination of individual specific dynamics, lifestyle, and other factors. An extension of this work will include testing personalized Blood Glucose prediction models in a more challenging situation involving real subjects.

REFERENCES:

1. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care, Vol. 31, No. 1, 2008, 55-60, 1935- 5548.
2. I. Eleni Georga, C. Vasilios Protopappas and I. DimitriosFotiadi,. "Glucose Prediction in Type 1 and Type 2 Diabetic Patients Using Data Driven Techniques, "Knowledge-Oriented Applications in Data Mining, In Tech pp 277-296, 2011.
3. V. Tresp, T. Briegel, and J. Moody, "Neural-network models for the blood glucose metabolism of a diabetic," IEEETransactions on Neural Networks, Vol. 10, No. 5, 1999, 1204-1213, 1045-9227.
4. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006, New York
5. S. Haykin, Neural networks and learning machines. Pearson 2008.
6. W.D. Patterson, Artificial Neural Networks- Theory and Applications, PrenticeHall,Singapore.1996.

7. M. Pradhan and R. Sahu, "Predict the onset of diabetes disease using Artificial Neural Network (ANN)," International Journal of Computer Science & Emerging Technologies, 303Volume2,Issue2, April2011.
8. W, Sandham, D,Nikoletou, D.Hamilton, K, Paterson, A. Japp and C. MacGregor, "BLOOD Glucose Prediction for Diabetes THERAPY USING A RECURRENT Artificial Neural Networks ", EUSIPCO, Rhodes, 1998, PP. 673-676
9. M, Divya, R. Chhabra, S. Kaur and S., "Diabetes Detection Using Artificial Neural Networks & Back-Propagation Algorithm", International Journal of Scientific & Technology Research". V 2, ISSUE 1, JANUARY 2013.
10. G, Robertson, E. Lehmann, W. Sandham, and D. Hamilton Blood, "Glucose Prediction Using Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study". Journal of Electrical and Computer Engineering, V 2011 (2011), Article ID 681786, 11 pages.
11. Van Cauter E, Shapiro E, Tillil H, Polonsky K. Circadian modulation of glucose and insulin responses to meals: relationship to cortisol rhythm. AmJ Phys. 1992;262:E467–E475.
12. Gogou G, Maglaveras N, Ambrosiadou BV, Goulis D, Pappas C. A neural network approach in diabetes management by insulin administration. J Med Syst. 2001;25(2):119–131.
13. Karim EJ. Neural network control of type I diabetes mellitus. Bioprocess Biosyst Eng. 2005;27(2):71–76. 23. Pender JE. Modelling of blood glucose levels using artificial neural networks. Dissertation. Glasgow, Scotland: University of Strathclyde; 1997.
14. Sandham WA, Nikoletou D, Hamilton DJ, Patterson K, Japp A, Macgregor C. Blood glucose prediction for diabetes therapy using a recurrent artificial neural network. Conf Proc EUSIPCO. 1998; (11):673–676.
15. Ibrahim, S. Jafar Ali, and M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with hepatocellular carcinoma based on multi-source bat algorithm based random

walk." Measurement 141(2019):176-183.
https://doi.org/10.1016/j.measurement.2019.02.056

16. Thangamani, M., and S. Jafar Ali Ibrahim. "Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso-Wc) and Gene Ontology for Microarray Gene Expression."In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, pp. 48-55. 2018 .https://dl.acm.org/doi/abs/10.1145/3299852.3299866

Plant Species Identification And Disease Detection Using Deep Learning

Nannuri Suresh [1], M. Nandini [2], B. Tejaswini [3], CH. Srikanth [4], SD. Arief [5]

Assistant Professor (IT), Dept. of Information Technology, Qis College of Engineering and Technology,Andhra Pradesh, India (1)

Final Year (B. Tech), Dept. of Information Technology, Qis College of Engineering and Technology,Andhra Pradesh, India (2,3,4,5)

ABSTRACT:

Plant diseases affect the growth of the species and their respective crops, early identification of the plant disease prevents the losses in the yield and improves the quality and quantity of the agriculture products. Many machine learning models are used to detect the disease in the plant but, after the advancements of deep learning models, this area of research appears to have a great potential in terms of increasing accuracy.

The proposed system identifies the plant species and disease of the leaf. The dataset we got from internet Kaggle is segregated and the different plant species are identified and are renamed to form a proper database then obtain test-database which consists of various plant diseases that are used for checking the accuracy of the project. Then using training data, we will train our classifier and then output will be predicted with better accuracy, we used google Net model to train the data, which consists of different layers which are used for predicting the disease.

The existing system the farmers are using for the detection of diseases in the plants is that- they could be identified through the naked eye and their knowledge about plant disease. For doing so, on large number of plants is time consuming, difficult and accuracy is not good, Consulting experts is of great cost. In such kind of conditions to improve the accuracy rate and make it more beneficial suggested techniques are implemented for the detection of the diseases that makes the process cheaper and easier This review provides a comprehensive explanation of deep learning models used to visualize various

plant diseases. In addition, some research gaps are identified from which to obtain greater transparency for detecting diseases in plants, even before their symptoms appear clearly.

1. INTRODUCTION

An automated plant species identification and disease diagnosis system could help botanists and layman in identifying plant species rapidly. Deep learning is robust for feature extraction as it is superior in pro- viding deeper information of images. In this research, the leaf images were pre-processed and the features were extracted by using Convolutional Neural Network (CNN). This paper proposes a mathematical model of plant disease detection and recognition based on deep learning, which improves accuracy, generality, and training efficiency. The segmented leaves are input into the transfer learning model and trained by the dataset of diseased leaves under simple background.

2. OBJECTIVE

Aim of this project is to develop a system which identifies the species and disease of a plant by using Google Net architecture of CNN algorithm, project is to develop a system which can perform early prediction of disease in plants with a higher accuracy Image processing is also used to represent the output in the forms of images and charts.

The image processing could be used in the field of agriculture for several applications. It includes detection of diseased leaf, stem or fruit, to measure the affected area by disease, to determine the color of the affected area. Tomato cultivation is one of the most remunerative farming enterprises in India. The naked eye observation by the experts is approach usually taken in identification and detection of plants. This approach is time consuming in huge farms or land areas. The use of image processing

3. EXISTING SYSTEM

The existing method for plant disease detection is simply naked eye observation by experts through which identification and detection of plant diseases is done. For doing so, a large team of experts as well as continuous monitoring of plant is required, which costs very high when we do with large farms.

4. PROPOSED SYSTEM

Proposed system opted to develop an Android application that detects plant diseases. It has thealgorithms and models to recognize species anddiseases in the crop leaves by usingConvolutional Neural Net- work. Proposed system uses Collab to edit source code. A dataset of 54,305 images of diseased and healthy plant leaves collected under controlled conditions Plant Village dataset. The images cover 14 species of plants, including: apple, grape, orange, pepper, potato, brinjal, lady's finger, and tomato.

It contains images of 17 basic diseases, 4 bacterial diseases, 2 diseases caused by mold (oomycete), 2 viral diseases and 1 disease caused by a mite. 12 crop species also have healthy leaf images that are not visibly affected by disease. Our dataset contains solutions for several plant textures such as.

6. EXPERIMENTAL STUDY

Techniques in detection and identification of tomato plant diseases in the earlier stages and thereby the quality of the product could be increased. These systems monitor the plant such as leaves and stem and any variation observed from its characteristic features, variation will be automatically identified and also will be informed to the user.

5.LITERATURE SURVEY

Paper [1] presents classification and detection techniques that can be used for plant leaf disease classification. Here preprocess is done before feature extraction. RGB images are converted into white and then converted into grey level image to extract the image of vein from each leaf. Then basic Morphological functions are applied on the image. Then the image is converted

into binary image. After that if binary pixel value is 0 its converted to corresponding RGB image value. Finally, by using Pearson correlation and Dominatingfeature set and Naive Bayesian classifierdisease is detected.

In paper [2] there are four steps. Out of them the first one is gathering image from several part of the country for training and testing. Second part is applying Gaussian filter is used to remove all the noise and thresholding is done to get the all- green color component. K-means clustering is used for segmentation. All RGB images are converted into HSV for extracting feature.

The paper [3] presents the technique of detecting jute plant disease using image processing. Image is captured and then it is realized to match the size of the image to be stored in the database. Then the image is enhanced in quality and noises are removed. Hue based segmentation is applied on the image with customized thresholding formula.

6.1. FEASABILITY STUDY

In this review, we present a comprehensive and critical survey on image-based plant leaf disease prediction techniques. Diseases in plants cause major production and economic losses in agricultural industry worldwide. Monitoring of health and detection of diseases in plants and trees is critical for sustainable agriculture. To the best of our knowledge, there is no sensor commercially available for real- time assessment of health conditions in trees. Currently, scouting is most widely used mechanism for monitoring stress in trees, which is an expensive, labor-intensive, and time- consuming process. Molecular techniques such as polymerase chain reaction are used for the identification of plant diseases that require detailed sampling and processing procedure. Early information on crop health and disease detection can facilitate the control of diseases through proper management strategies such as vector control through pesticide applications, fungicide applications, and disease-specific chemical applications; and can improve productivity. The aim of this research is to propose and evaluate a framework for detection of plant leaf diseases. Studies show that re- lying on pure naked-eye observation of experts to detect such diseases can be prohibitively expensive, especially in developing countries.

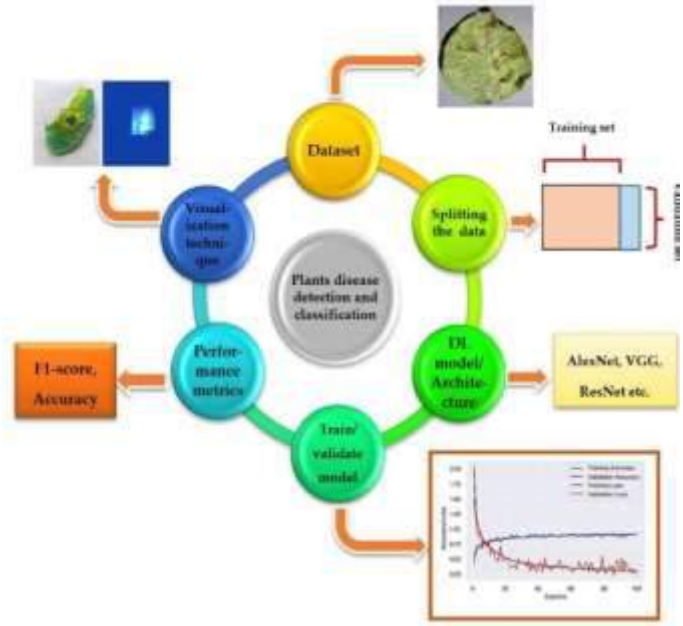
6.1.1 Economic Feasibility

Agricultural productivity is something on which economy highly depends. This is the oneof the reasons that disease detection in plants plays an important role in agriculture field, as having disease in plants are quite natural. If proper care is not taken in this area, then it causes serious effects on plants and due to which respective product quality, quantity or productivity is affected. For instance, a disease named little leaf disease is a hazardous disease found in pine trees in United States. Detection of plant disease through some automatic technique is beneficial as it reduces a largework of monitoring in big farms of crops, and at very early stage itself it detects the symptomsof diseases i.e., when they appear on plant leaves. This paper presents an algorithm for image segmentation technique which is used for automatic detection and classification of plant leaf diseases. It also covers survey on different diseases classification techniques that can be used for plant leaf disease detection. Imagesegmentation, which is an important aspect for disease detection in plant leaf disease, is done by using genetic algorithm.

6.1.2 Technical Feasibility

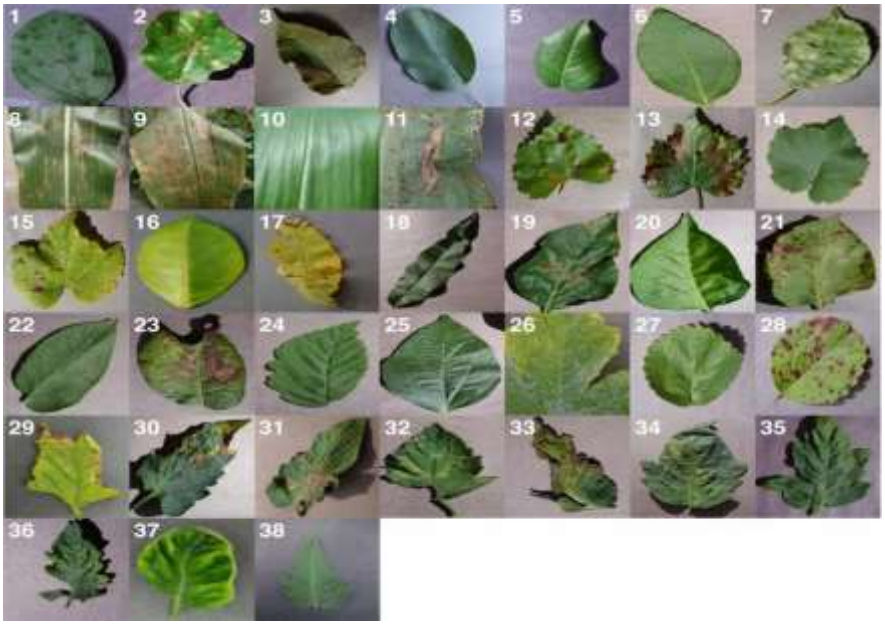
The focus is on enhancing productivity, without considering the ecological impacts that has resulted in environmental degradation. As dis- ease of the plants is inevitable, detecting disease plays a major role in the field of agriculture. Plant pathogens consist of fungi, organism, bacteria, viruses, phytoplasmas, viroid's etc. Three components are absolutely necessary for diseases to occur in any plant system and which may infect all types of plant tissues including leaves, shoots, stems, crowns, roots, tuber, fruits, seeds and vascular tissues. Therefore, detection and classification of diseases is an important and urgent task. The necked eye observation of experts is the main approach adopted in practice for detection and identification of plant diseases. How- ever, this requires continuous monitoring of experts which might be prohibitively expensive in large farms. We can analyze the image of disease leaves by using computer image processing technology and ex- tract the features of disease spot according to color, texture and other characteristics from a quantitative point of view.

6.2 ARCHITECTURE



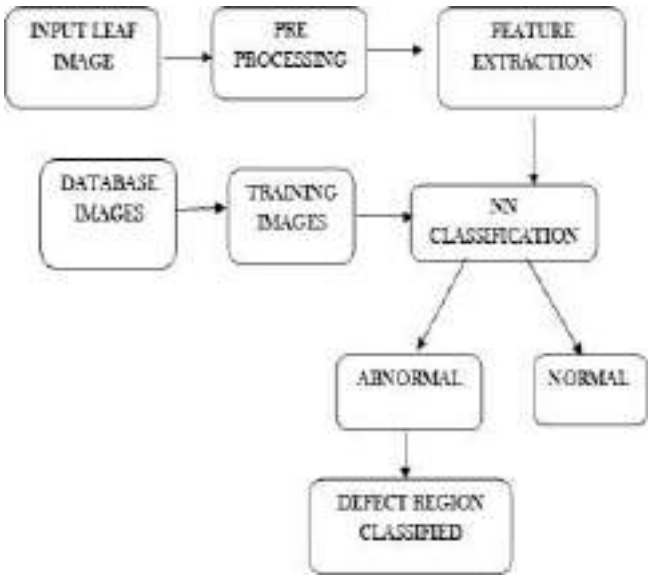
6.3 METHODOLOGY

To get a sense of how our approaches willperform on new unseen data, and also to keep atrack of if any of our approaches are overfit- ting, we run all our experiments across a wholerange of train-test set splits , namely 80–20 , the Plant Village dataset has multiple images of thesame leaf (taken from different orientations), and we have the mappings of such cases for 41,112 images out of the 54,306 images; and during all these test-train splits, we make sure all the images of the same leaf goes either in thetraining set or the testing set. Further, for every experiment, we compute the mean precision, mean recall, mean F1 score, along with the overall accuracy over the whole period of training at regular intervals (at the end of everyepoch). We use the final mean F1 score for thecomparison of results across all of the differentexperimental configurations. We evaluate the applicability of deep convolutional neural networks for the classification problem described above. We focus on popular CNN architecture Google Net. The Google Net architecture on the other hand is a much deeperand wider architecture with 22 layers, while stillhaving considerably lower number of parameters (5 million parameters) in the network than Alex Net (60 million parameters).



6.4 DESCRIPTION Data flow diagram:

The testing and training dataset are used in CNN model to predict the leaf disease the last level comprises of both CNN and dense CNN model. It is used to gain more accuracy.



6.5 Results

The plant disease recognition model based on deep learning has the characteristics of unsupervised, high accuracy, good universality, and high training efficiency. However, there aremany challenges in accuracy practicability of plant disease detection in the complex environment. In order to solve these problems and optimize the identification method, this paper proposes a recognition model integratingRPN algorithm, CV algorithm, and TL algorithm, which can effectively solve the problem of plant disease identification in the complex environment. The model not only adapts to complex environments, but also increases the accuracy of identification. Compared with the traditional model, the modelproposed in this paper not only guarantees the robustness of the convolutional neural network, but also reduces the number and quality requirements of the convolutional neural network on the data set and obtains better results. Therefore, the model could help agricultural production personnel to prevent and cure the plant disease quickly.

6.6 CONCLUSION

In conclusion, this research is about plant species identification and disease detection by using deep learning via framework TensorFlow. It has three (3) objectives that have achieved throughout this research. The objectives are linked directly with conclusions because it can determine whether all objectives are successfully achieved or not. It can be concluded that all results that have been obtained, showed quite impressive outcomes. Implementation of deep learning by using framework TensorFlow also gave good results as it is able to simulate, train and classified with up to 90% percent of accuracy towards differentplants that have become a trained model. Lastly, Python have been used as the programming language throughout this research since it comes together with framework TensorFlow which leads to designing of the system involved Python from start until it ends

REFERENCES

1. Wen J, Shi Y, Zhou X, Xue Y. Crop disease classification on inadequate low- resolution target images.Sensors.2020;20(16):4601.
2. Atila M, Uar M, Akyol K, Uar E. Plant leaf disease classification using efficientnet deep learning model. Ecol Inform. 2021; 61:101182.
3. Wiesner-Hanks T, Wu H, Stewart E, Dechant C, Nelson RJ. Millimeter-level plant disease detection from aerial photographs via deep learning and crowd- sourced data. Front Plant Sci. 2019;10:1550
4. Barbedo JG. Plant disease identification from individual lesions and spots using deep learning. Biosyst Eng. 2019; 180:96–107. Issue, Pageno.
5. Y. Ampatzidis, L. De Bellis, and A. Luvisi,"I Pathology: robotic applications and management of plants and plant diseases,"*Sustainability*, vol. 9, no. 6, p. 1010, 2017.
6. A. Breuker's, D. L. Kettenis, M. Mourits,
7. W. V. D. Werf, and A. O. Lansink, "Individual-based models in the analysis of disease transmission in plant production chains: an application to potato brown rot," *Academy of Sciences*, vol. 90, no. 1–3, pp. 112–131, 2006.
8. S. Ghosal, D. Blystone, A. K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, "An explainable deep machinevision framework for plant stress phenotyping," *Proceedings of the NationalAcademy of Sciences*, vol. 115, no. 18, pp. 4613–4618, 2018.
9. E.-C. Oerke, "Crop losses to pests," *TheJournal of Agricultural Science*, vol. 144, no.1, pp. 31–43, 2006.
10. X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species throughimage features analysis and One Class Classifiers," *Computers and Electronics in Agriculture*, vol. 156, pp. 96–104, 2019.
11. J. G. A. Barbedo, "Factors influencing the use of deep learning for plant

disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, 2018.

12. G. Geetha Ramani and J. Arun Pandian,"Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.
13. P. F. Konstantinos, "Deep learning models for plant disease detection and diagnosis," *Computers & ElectricalEngineering*, vol. 145, pp. 311–318, 2018

Detection of Cyber Attacks in Network Using Machine Learning Techniques

Mr. N. SaiKiran[1], P. D. S. Pradeep Naidu[2], K. Harshini[3], M. Venkateswarlu[4], V. Surya Narayana Reddy[5]

[1]Associate Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Abstract- The development of effective techniques has been an urgent demand in the field of the cyber security community. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analysing the information for possible security problems. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively.

Keywords- CNN, LSTM, [[[image captioning, deep learning.

1. INTRODUCTION:

Cyber-crime is proliferating everywhere exploiting every kind of vulnerability to the computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of innovations gives incredible advantages to

people, organizations, and governments, be that as it may, messes come up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information, and so forth. Contingent upon these issues, digital fear-based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and national security by different gatherings, for example, criminal association, proficient people, and digital activists. Along these lines, Intrusion Detection Systems (IDS) have been created to maintain a strategic distance from digital assaults. Right now, learning the bolster support vector machine (SVM) calculations were utilized to recognize port sweep endeavors dependent on the new CICIDS2017 dataset with 97.80%, 69.79% precision rates were accomplished individually. Rather than SVM we can introduce some other algorithms like the random forest, CNN, ANN where these algorithms can acquire accuracies like SVM – 93.29, CNN – 63.52, Random Forest – 99.93, ANN – 99.11.

LITERATURE SURVEY:

Ate of attacks against networked systems has increased melodramatically, and the strategies used by the attackers are continuing to evolve.For example, the privacy of important information, security of stored data platforms, availability of knowledge, etc. Depending on these problems, cyber terrorism is one of the most important issues in today's world. Cyber terror, which caused a lot of problems to individuals and institutions, has reached a level that could threaten public and country security by various groups such as criminal organizations, professional persons, and cyber activists. Intrusion detection is one of the solutions to these attacks. A free and effective approach for designing Intrusion Detection Systems (IDS) is Machine Learning. In this study, deep learning and support vector machine (SVM) algorithms were used to detect port scan attempts based on the new CICIDS2017 dataset Introduction Network Intrusion Detection System (IDS) is a software-based application or a hardware device that is used to identify malicious behavior in the network [1,2]. Based on the detection technique,

intrusion detection is classified into anomaly-based and signature-based. IDS developers employ various techniques for intrusion detection. Information security is the process of protecting 12 information from unauthorized access, usage, disclosure, destruction, modification, or damage. The terms ”Information security”, computer security” and information insurance” are often used interchangeably.

EXISTING SYSTEM:

`Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues.

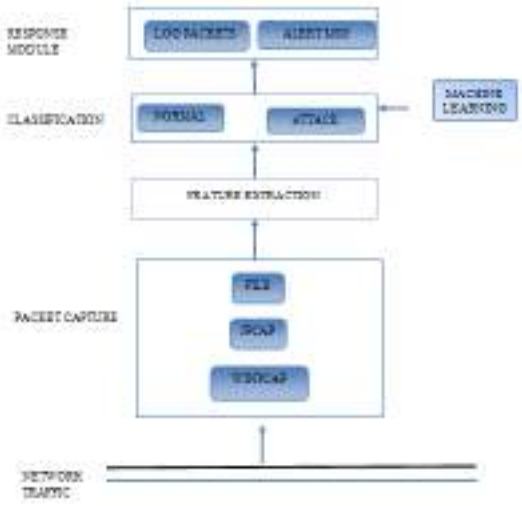
An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent to keep our data secure. One significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats. If they don't take care to monitor the false positives, real attacks can slip through or be ignored.

When an IDS detects suspicious activity, the violation is typically reported to a security information and event management (SIEM) system where real threats are ultimately determined amid benign traffic abnormalities or other false alarms. However, the longer it takes to distinguish a threat, the more damage can be done. An IDS is immensely helpful for monitoring the network, but their usefulness all depends on what you do with the information that they give you. Because detection tools don't block or resolve potential issues, they are ineffective at adding a layer of security unless you have the right personnel and policy to administer them and act on any threats. An IDS cannot see into encrypted packets, so intruders can use them to slip into the network.

An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent

to keep our data secure. One significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats. An IDS can be tuned to reduce the number of false positives; however, your engineers will still have to spend time responding to them. If they don't take care to monitor the false positives, real attacks can slip through or be ignored

SYSTEM ARCHITECTURE:



Network Traffic:

Network traffic refers to the amount of data moving across a network at a given point of time. Network data is mostly encapsulated in network packets, which provide the load in the network. Network traffic is the main component for network traffic measurement, network traffic control and simulation. The Proper analysis of network traffic provides the organization with the following benefits: Identifying network bottlenecks - There could be users or applications that consume high amounts of bandwidth, thus constituting a major part of the network traffic. Different solutions can be implemented to tackle these. Network security - Unusual amount of traffic in a network is a possible sign of an attack. Network traffic reports provide valuable insights into preventing such attacks. Network engineering - Knowing the usage levels of the network allows future requirements to be analysed.

Packet capture : Packet Capture is a networking term for intercepting a data packet that is crossing a specific point in a data network. Once a packet is captured in real-time, it is stored for a period of time so that it can be analysed, and then either be downloaded, archived or discarded. Packets are captured and examined to help diagnose and solve network problems such as: Identifying security threats Troubleshooting undesirable network behaviours Identifying network congestion Identifying data/packet loss Forensic network analysis.

Classification : Classification is another extensively used supervisory machine learning task. In cyber security, spam detection is successfully implemented by ML based classifiers which involves discriminating a given email message as spam or not. The spam filter models are able to separate spam messages from non-spam messages. Machine learning techniques for classification include Logistic Regression, K-Nearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest Classification

IMPLEMENTATION DETAILS:

The system is implemented by using ANACONDA software , Anaconda is the world's most popular data science platform and the foundation of modern machine learning

Securing confidential information Algorithms:

Artificial Neural Network (ANN):

The plan thought of an ANN is to mirror the manner in which human cere brum work. An ANN contains an info layer, a few secret layers and a yield layer. The units in neighbouring layers are completely associated. An ANN contains a colossal number of units and can hypothetically estimated subjective capacities; subs equently, it has solid fitting capacity, particularly for nonlinear capacities. Because of the perplexing model design, preparing ANNs is tedious.

Support Vector Machine (SVM):

The system in SVMs is to discover a maximum edge partition hyperplane in the n-measurement highlight space. SVMs can accomplish satisfying outcomes even with limited scope preparing sets in light of the fact that the partition hyperplane is resolved simply by few help vectors. In any case, SVMs are delicate to commotion close the hyperplane.

K-Nearest Neighbour (KNN).

The centre thought of KNN depends on the complex theory. On the off chance that the majority of an example's neighbours have a place with a similar class, the example has a high likelihood of having a place with the class. In this manner, the grouping result issimply identified with the top-k closest neighbours. The boundary k enormously impacts the presentation of KNN models. The more modest k is, the more intricate the modelis and the higher the danger of overfitting. On the other hand ,the bigger k is, the easier the model is and the more fragile the fitting capacity.

PROPOSED SYSTEM:

Machine Learning algorithms can be used to train and detect if there has been a cyber attack. As soon as the attack is detected, an email notification can be sent to the security engineers or users. Any classification algorithm can be used to categorize if it is a DOS/DDOS attack o[r not. One example of a classification algorithm is Support Vector Machine (SVM) which is a supervised learning method that analyses data and recognizes patterns. Since we cannot control when, where or how an attack may come our way, and absolute[[e prevention against these cannot be guaranteed yet, our best shot for now is early detection which will help mitigate the risk of irreparable damage such incidents can cause. Organizations can use existing solutions or build their own to detect cyber attacks at a very early stage to minimize the impact. Any system that requires minimal human intervention would be ideal.

Problem Modelling Network admins The following steps are the functions of Network admin: Intercept network traffic. Read and store the data packets information. Check for alerts regarding the cyber-attacks and network.

SYSTEM ANALYSIS:

a) **System**: A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objective.

b) **System Analysis**: System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls, feedback and environment.

IMPLEMENTATION:

The system is implemented by using ANACONDA software , Anaconda is the world's most popular data science platform and the foundation of modern machine learning.

Securing confidential information Algorithms:

Artificial Neural Network (ANN):

The plan thought of an ANN is to mirror the manner in which human cerebrum work. An ANN contains an info layer, a few secret layers and a yield layer. The units in neighburing layers are completely associated. An ANN contains a colossal number of units and can hypothetically estimated subjective capacities; subs equently, it has solid fitting capacity, particularly for nonlinear capacities. Because of the perplexing model design, preparing ANNs is tedious.

Support Vector Machine (SVM):

The system in SVMs is to discover a maximum edge partition hyperplane in the n-measurement highlight space. SVMs can accomplish satisfying outcomes even with limited scope preparing sets in light of the fact that the partition hyperplane is resolved simply by few help vectors. In any case, SVMs are delicate to commotion close the hyperplane

K-Nearest Neighbor (KNN):

The center thought of KNN depends on the complex theory. On the off chance that the majority of an example's neighbors have a place with a similar class, the example has a high likelihood of having a place with the class. In this manner, the grouping result issimply identified with the top-kclosest neighbors. The boundary k enormously impacts the [presentation of KNN models. The more modest k is, the more intricate the model is and the higher the danger of over fitting. On the other hand ,the bigger k is, the easier the model is and the more fragile the fitting capacity.

SOFTWARE REQUIREMENTS:

Processor : Intel(R)Core (TM)I5
RAM : 2.00GB
System Type : 64Bit Operating System.

HARDWARE REQUIREMENTS:

Python 3.8.3
Visual studio code
Python Django Web framework
Beautiful Soup.

FUNDAMENTALS OF MACHINE LEARNING:

Support vector machine (SVM) is another widely used supervised machine learning model. SVM works to find hyper plane with most suitable dataset distribution by classifying the data into two classes on both sides of the hyperplane. Both sides of the hyperplane donate a separate class. The class of

every data point depends on the side of the hyperplane it lands. Support vector machine has a high consumption of space and time to handle larger and noisier datasets . The computational complexity of SVM is O(n2) where n represents the number of instances . A matrix that is used to evaluate the performance of machine learning classifier is called a confusion matrix .

Cyber attack detection techniques fall into two categories: signature-based and inconsistency-based. In both cases, machine learning techniques are used. The authors of improved the detection of Denial-of-Service (DoS) attacks. The Naive Bayes classifier was created based on the element vectors, which included different User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) bundles and their sizes. It has also been shown that Discrete Wavelet Transform and Matching Pursuit may successfully be used to calculate highlights depending on various organisational boundaries.

Error Rate:

The error rate (E Rate) is a percentage of the total number of misclassified instances to all instances of the dataset.

E Rate = (F Positive + F Negative)/ (T Negative + F Positive + F Negative + T Positive)

Recall:

The recall is a percentage of correctly classified positive instances to the total number of positive instances classified in the dataset.

Recall = T Positive / (T Positive + F Negative)

Precision:

The precision is a percentage of the total number of positive instances classified to the total number of positive instances.

Precision= T Positive / (T Positive + F Positive)

MODULES:

- Classification
- Feature extraction

- Detection
- Evaluation

METHODOLOGY:

- Support vector machine (SVM)
- It is a classification method.

In this algorithm, we plot each data item as a point in n-dimensional space (where n is number offeatures you have) with the value of each feature being the value of a particular coordinate.

RESULT:

The experiments were conducted in Machine learning libraries like numpy , pandas, scik [it learn. Python language is used to develop the application with jupyter notebook IDE .Predictions can be done by four algorithms like SVM,ANN, RF, CNN this paper helps to identify which algorithm predicts the best accuracy rates which helps to predict best results to identify the cyber attacks happened or not. Fig: 2 Protocol Type Distribution.

CONCLUSION:

At the present time, assessments of help vector machine ,ANN, CNN, Random Forest and significant learning estimations reliant upon current CICIDS2017datasetwere presented moderately. Results show that the significa ntlearning estimation performed generally best results over SVM, ANN, RFand CNN. We will use port scope attempts just as other attack types with AI and significant learning computations, a pache Hadoop and shimmer advancements together ward on this dataset later on. Every one of these estimation assists us with recognizing the digital assault in network. It occurs in the manner that when we think about long back a long time there might be such countless assaults occurred so when these assaults are perceived then the highlights at which esteems these assaults are going on will be put away

in some datasets. So by utilizing these datasets we will anticipate if digital assault is finished. These forecasts should be possible by four calculations like SVM,ANN, RF, CNN this paper assists with distinguishing which calculation predicts the best precision rates which assists with foreseeing best outcomes to recognize the digital assaults occurred or not.

REFERENCES:

1. K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.
2. R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.
3. M. Baykara, R. Das¸, and I. Karadogan, "Bilgi g ¨uv enligis is tem lerindekullanilanarac¸ larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp.231– 239.
4. RashmiT V. "Predicting the System Failures Using Machine LearningAlgorithms".International Journal of Advanced Scientific Innovation,vol. 1, no. 1, Dec. 2020, doi:10.5281/zenodo.4641686.
5. S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1.IEEE, 2003, pp. 130– 138.
6. K. Ibrahimi and M. Ouaddane, "Management of intrusion detectionsystems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1
7. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong,pp352357http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf

8. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IoT Sensor Networks."Advances in Information Communication Technology and Computing, pp. 461-471. Springer, Singapore, 2022.https://link.springer.com/chapter/10.1007/978-981-19-0619-0_41

Effective Heart Disease Prediction Using Hybrid Machine Learning Technique

P. Kiran[1],A. Swathi[2],M. Sindhu[3],Y. Manikanta[4], K.Mahesh Babu[5]

Assistant Professor (IT), Dept. of Information Technology, QIS College of Engineering & Technology, India(1).

Final Year (B.Tech), Dept. of Information Technology, QIS College of Engineering & Technology, India(2,3,4,5)

Abstract:

Heart disease is one of the most significant causesof mortality in the world today. Prediction of cardiovascular.Disease is a critical challenge in the area of clinical dataanalysisIn this project, we propose a novel method that aims at finding significant features by applying machine learning techniques (EML) resulting in improving the accuracy in theprediction of Heart disease.HRFLM (Hybrid Random Forest Linear Model) Technique proved to be quite accurate in the prediction of heart disease. by using entropy feature selection technique and removing unnecessary features, different classification techniques such that Gaussian Naïve Bayes, Support Vector Machine, Hybrid Random Forest with Linear Model, and Extension extreme Machine Learning Technique are used on heart disease dataset for better prediction. Different performance measurement factors such as accuracy precision, recall, sensitivity, specificity, and F1-score are considered to determine the performance of the classification techniques. Our project compares the performances of the classification algorithms in the prediction of heart disease. It tries to find out the best classifier for the detection of heart diseases.

Introduction:

1. Background

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body's cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood

vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart diseases which are commonly known as cardiovascular diseases (CVD). Heart diseases are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middleincome and low-income countries.

Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack. Therefore, detection of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases. Due to the development of advance healthcare systems, lots of patient data are nowadays available (i.e. Big Data in Electronic Health Record System) which can be used for designing predictive models for Cardiovascular diseases.

Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. "Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data". Nowadays, a huge amount of data pertaining to disease diagnosis, patients etc. are generated by healthcare industries. Data mining provides a number of techniques which discover hidden patterns or similarities from data. Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets.

LITERATURE SURVEY:

In year 2000, research conducted by Shusaku Tsumoto says that as we human beings are unable to arrange data if it is huge in size we should use the data mining techniques that are available for finding different patterns from the available huge database and can be used again for clinical research and perform various operations on it. Y. Alp Aslandogan, et. al. (2004), worked on three

different classifiers called K-nearest Neighbour (KNN), Decision Tree, Naïve Bayesian and used Dempsters' rule for this three viewpoint to appear as one concluding decision. This classification based on the combined idea show increased accuracy. Carlos Ordonez (2004), Assessed the problematic to recognize and forecast the rule of relationship for the heart disease. A dataset involving medical history of the patients having heart disease with the aspects of risk factors was accessed by him, measurements of narrowed artery and heart perfusion. All these restrictions were announced to shrink the digit of designs, these are as follows: 1) The features should seem on a single side of the rule. 2) The rule should distinct various features into the different groups. 3) The count of features available from the rule is organized by medical history of people having heart disease only. The occurrence or the nonappearance of heart disease was predicted by the author in four heart veins with the two clusters of rules. Franck Le Duff (2004), worked on creating Decision tree quickly with clinical data of the physician or service. He suggested few data mining techniques which can help cardiologists in the predication survival of patients. The main drawback of the system was that the user needs to have knowledge of the techniques and we should collect sufficient data for creating an suitable model. Boleslaw Szymanski, et. al. (2006), operated on a novel experiential to check the aptitude of calculation of scarce kernel in SUPANOVA. The author used this technique on a standard boston housing market dataset for discovering heart diseases, measurement of heart activities and prediction of heart diseases were found 83.7% correct which were measured with the help of support vector machine and kernel equivalent to it. A quality result is gained by spline kernel with the help of standard boston housing market database.

Kiyong Noh, et. al. (2006) made use of a classification technique for removal of multiparametric structures by accessing HRV and ECG signals. Kiyong used the FP growth algorithm as the foundation of this technique that is associative. A rule consistency degree was gained which allows a robust press on trimming designs in the method of producing designs. HeonGyu Lee, et. al. (2007), operated for the operation systems of Arithmetical and cataloguing for the addition chief of the multi-parametric feature through direct and nonlinear

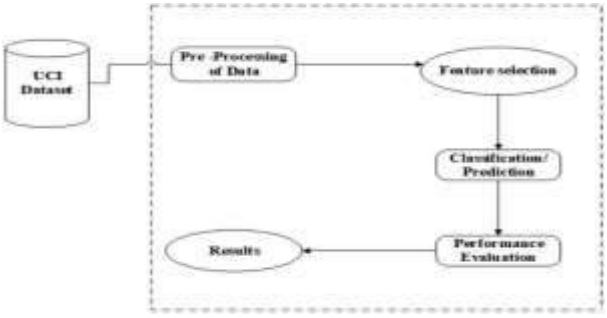
features of Heart Rate Variability (HRV). The dissimilar classifiers existing are cataloguing grounded on Decision Tree (C4.5), Multiple Association Rules (CMAR) and Bayesian classifiers, and Support Vector Machine (SVM) that are investigated for the valuation of the linear and nonlinear features of the HRV tables .Niti Guru, et. al. (2007), functioned for forecasting of heart disease, Blood Stress and Sugar by the aid of neural systems. Hearings were accepted out on example best ever of patients. The neural system is verified with 13 types, as blood pressure, period, angiography etc. Controlled network was used for analysis of heart diseases. Training was accepted out with the support of a back-propagation technique. The secretive data was nourished at certain times by the doctor; the acknowledged technique applied on the unidentified data since the judgments with trained data and caused a grade of possible ailments that the patient is inclining to heart disease. Hai Wang, et al. (2008), deliberated the part of medicinal experts in medical data mining also on obtaining a model for medical awareness achievement using data mining. Sellappan Palaniappan, etal. (2008), industrialized IHDPS-Intelligent Heart Disease Prediction System by means of data mining algorithm, i.e. Naïve Bayes, Decision Trees and Neural Network. Each process has its own authority to advance right results. The unknown designs and association amongst them have were used to paradigm this method. IHDPS is web-based user-friendly, mountable, trustworthy and stretchy and justifiable Latha Parthiban, et. al. (2008), operated on the foundation of CANFIS (co-active neuro - fuzzy implication method) for identification of heart disease. CANFIS model established the disease by integrating the neural network and fuzzy logic methods and later combined with the genetic algorithm. On the grounds of the training presentations and classification correctness found, the performance of the CANFIS model were estimated. The CANFIS prototypical is exposed as the possible for estimation of heart disease. Chaitrali S. D., (2012), investigated a computation structures for heart syndrome with the help of full amount of input characteristics. A few terms related to medical like blood pressure, sex, cholesterol and 13 more attributes like this were recycled to predict the heart disease to a particular person or patient. He also made use of two different attributes like smoking and obesity.

SYSTEM ARCHITECTURE:

System design is the process of defining elements of a system like modules, architecture, components and their interface and data for a system based on the specified requirements. The SystemDesignDocument describes the systemrequirements,operatingenvironment, systemand subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces.

It is a separable component, frequently one that is interchangeable with others, for assembly into units ofdiffering size, complexityor function. This section describes the system in narrative form using non-technical terms. It should provide a high-level system architecture diagram showing a subsystem breakout of the system, if applicable. The high- level system architecture or subsystem diagrams should, if applicable, show interfaces to external systems.

4.1 System Architecture:



Objective Function:

1. **UCI Dataset:** The **UCI** Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. Here the Heart Disease Data set from the Repository is taken.

2. **Pre-processing of Data:** Data Preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain

behaviors or trends, and is likely to contain many errors. Data Preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing. Data preprocessing is used in database-driven applications such as Customer relationship Management and rule-based applications.

3. **Feature Selection :** Feature selection refers to the process of reducing the inputs for processing and analysis, or of finding the most meaningful inputs. A related term,feature engineering (or feature extraction), refers to the process of extracting useful information or features from existing data

4. **Classification / Prediction: Classification** is a task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. An easy to understand example is classifying emails as "spam" or "not spam". **Classification predictive** Modelling involves assigning a class label to input examples.

5. **Performance Evaluation :** An evaluation metric quantifies the performance of a predictive model. This typically involves training a modelon a dataset, using the model to make predictions on a holdout dataset not used during training, then comparing the predictions to the expected values in the holdout dataset. The choice of evaluation metrics depends on a given machine learning task (such as classification, regression, ranking, clustering, topic modeling, among others). Some metrics, such as precision-recall, are useful for multiple tasks.

4.2 UML Diagrams:

UML diagrams represent static and dynamic views of a system model. The static view includes class diagrams and composite structure diagrams, which emphasize static structure of systems using objects, attributes, operations and relations.

UML (Unified Modeling Language) is a standard vernacular for choosing, envisioning, making, and specifying the collectibles of programming

structures. UML is a pictorial vernacular used to make programming blue prints. It is in like way used to exhibit non programming structures similarly like process stream in a gathering unit and so forth.

GOALS :

The Primary goals in the design of the UML are as follows:

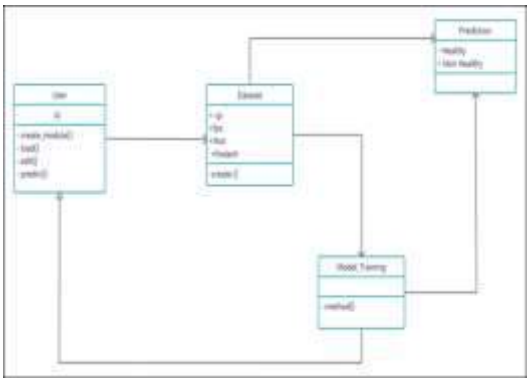
- Provide users a ready-to-use, expressive visual modeling language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Provide a formal basis for understanding the modeling language.
- Encourage the growth of OO tools market.

Diagram plays a very important role in the UML. These are kinds of modeling diagrams are as follows:

- Class Diagram
- Use case Diagram
- Sequence Diagram
- Activity Diagram
- Component Diagram
- Collaboration Diagram
- State Chart Diagram

4.2.1. CLASS DIAGRAM :

A class diagram is an illustration of the relationships and source code dependencies among classes in the Unified Modeling Language(UML) in the context, a class defines the methods and variables in an object, which is a specific entity in a program or the unit of code representing the entity. The class graph is the most normally pulled in layout UML. It addresses the static course of action Perspective of the structure. It solidifies the strategy of classes, interfaces, joint attempts and their affiliations.



ALGORITHM DESCRIPTION

5.2 Introduction to Python :

Below are some facts about Python.Python is currently the most widely used multi-purpose, high-level programming language.Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.

The biggest strength of Python is huge collection of standard library which can be used for the following –

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQtetc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

Features of Python:

Let's see how Python dominates over other languages.

1. Extensive Libraries:

Python downloads with an extensive library and it contain code for various purposes like regular expressions, documentation-generation, unit-

testing, web browsers, threading, databases, CGI, email, image manipulation, and more. So, we don't have to write the complete code for that manually.

2. Extensible:

As we have seen earlier, Python can be extended to other languages. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.

3. Embeddable:

Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add scripting capabilities to our code in the other language.

4. Improved Productivity:

The language's simplicity and extensive libraries render programmers more productive than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.

5. IOT Opportunities:

Since Python forms the basis of new platforms like Raspberry Pi, it finds the future bright for the Internet Of Things. This is a way to connect the language with the real world.

6. Simple and Easy:

When working with Java, you may have to create a class to print 'Hello World'. But in Python, just a print statement will do. It is also quite easy to learn, understand, and code. This is why when people pick up Python, they have a hard time adjusting to other more verbose languages like Java.

7. Readable:

Because it is not such a verbose language, reading Python is much like reading English. This is the reason why it is so easy to learn, understand, and code. It also does not need curly braces to define blocks, and indentation is mandatory. This further aids the readability of the code.

8. Object-Oriented:

This language supports both the procedural and object-oriented programming paradigms. While functions help us with code reusability, classes and objects let us model the real world. A class allows the encapsulation of data and functions into one.

9. Free and Open-Source:

Like we said earlier, Python is freely available. But not only can you download Python for free, but you can also download its source code, make changes to it, and even distribute it. It downloads with an extensive collection of libraries to help you with your tasks.

10. Portable:

When you code your project in a language like C++, you may need to make some changes to it if you want to run it on another platform. But it isn't the same with Python. Here, you need to code only once, and you can run it anywhere. This is called Write Once Run Anywhere (WORA). However, you need to be careful enough not to include any system-dependent features.

11. Interpreted:

Lastly, we will say that it is an interpreted language. Since statements are executed one by one, debugging is easier than in compiled languages.

5.1. Modules Module Description :

1) **Upload Module**: using this module wewill upload heart disease dataset of previous patients.
2) **Pre-process Module**: Using this module we will remove all those records which contains missing values. Dataset will be split into two parts called training and testing, all classifier will build train model using training data and then test train model by applying test data on that train model to get classification accuracy.
3) **SVM Module**: Using this module we will build train model using SVM algorithm and then apply test data on that SVM model to get classification accuracy.

4) **Naïve Bayes**: Using this module we will build train model by using Naïve Bayes algorithm and apply test data to get Naïve Bayes classification accuracy.
5) **HRFLM:** Propose Hybrid Algorithm which is combination of Linear model and Random Forest algorithm. Hybrid model will be generated by using both algorithms and then Voting classifier will be used to choose best performing algorithm.

Extension Extreme Machine Learning Module: This is an extra module which is built for extension purpose and this module is based on advance Extreme Machine Learning algorithm which can get better prediction accuracy compare to all algorithms. Extreme Learning Machine (ELM) is a novel method for pattern classification as well as function approximation. This method is essentially a single feed forward neural network; its structure consists of a s[ingle layer of hidden nodes, where the weights between inputs and hidden nodes are randomly assigned and remain constant during training and predicting phases. On the contrary, the weights that connect hidden nodes to outputs can be trained very fast. Experimental studies in the literature showed that ELMs can produce acceptable predictive performance and their computational cost is much lower than networks trained by the back-propagation algorithm.

6) Graph : This module display accuracy of all algorithms in graph format as comparison

Conclusion:

- In this paper, We introduce Heart disease prediction system with different classifier techniques and compared their performance.
- The techniques are SVM, Naive Bayes classifier, Regression, HRFLM and EML .
- Among them the proposed Hybrid HRFLM approach has produced higher accuracy level of 81.6% and Extension EML has produced accuracy level of 92.6% in prediction.

REFERENCES

[1] V. Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.

[2] Sellappan Palaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques", International Journal of Computer Science and Network Security, vol.8, no.8, pp. 343-350,2008.

[3] K.Srinivas, Dr.G.Ragavendra and Dr. A. Govardhan," A Survey on prediction of heart morbidity using data mining techniques",International Journal of Data Mining & Knowledge Management Process (IJDKP) vol.1, no.3, pp.14-34, May 2011.

[

A Machine Learning-Based Prediction Model for Preterm Birth in Rural Areas

Dr.S.Jafar Ali Ibrahim[1], Y. Prathyusha[2], B. Akhila[3], M. Rajini[4], U. Naveen Sai[5]

Associate Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Abstract:

Preterm birth (PTB) in a pregnant woman is the most serious issue in the field of Gynaecology and Obstetrics, especially in rural India. In recent years, various clinical prediction models for PTB have been developed to improve the accuracy of learning models. However, to the best of the authors' knowledge, most of them suffer from selecting the most accurate features from the medical dataset in linear time.The suggested approach is then applied on collected data to identify the excellent maternal features (text-based symptoms) present in pregnant women in order to classify all birth cases into term birth and PTB. The machine learning part of the model is implemented using three different classifiers, namely, decision tree (DT), logistic regression (LR), and support vector machine (SVM) for PTB prediction.

Introduction:

Preterm birth (PTB) is a serious public health problem that adversely affects both families and the society. It is a leading cause of neonatal mortality and morbidity across the world and also the second major cause of child deaths under the age of five years. Over the past two decades, PTB has been a significant research study in healthcare domain. Pregnancy and childbirth unlocked the door for medical experts and researchers to explore various effective strategies to reduce preterm birth in women having pregnancy-related complications. These strategies include healthcare services given to all pregnant women to control PTB and any medical interventions aimed to enhance the knowledge of women on early indications of pregnancy complications. The

maternal history of a pregnant woman is a key part of the neonatal studies for providing certain clinical treatments to newborn babies regarding their health, disease, care, and outcomes. Newborn babies are very special. They do not have any previous medical background, and their early neonatal path is directly connected to the maternal history of their mothers . The healthcare services also incorporate the arrangements of essential social and economic support for women before, during, and after pregnancy including educational, medical, and other training programs that facilitate healthy motherhood.

Background :

Preterm Birth (PTB): A Comprehensive Overview

Preterm or premature birth is defined as birth, for any reason, occurring before 37 completed weeks (or less than 259 days) of pregnancy. Every year, about fifteen million babies are born prematurely (before 37 completed weeks of gestation), and this is nearly equal to one-tenth of all babies around the world . According to the WHO reports studied in 2005, 12.9 million births or 9.6% of all births across the world occurred prematurely. The rate of preterm birth, however, significantly varies across the world. Preterm birth reflects the most prominent reasonfor neonatal morbidity and mortality. Categorization of PTB PTB can be classified into different categories based on gestational age at birth. The gestational age is defined as the time from the first day of the last menstrual period (LMP) of a woman to birth . The four categories of PTB are as follows:(i) Extreme PTB (under 28 Weeks). It is the birth that takes place before 28 weeks of pregnancy (ii)Very PTB (28 to 32 Weeks). It is the birth that takes place between 28 and 32 weeks of pregnancy (iii)Moderate PTB (32 to 34 Weeks). It is the birth that takes place between 32 and 34 weeks of pregnancy (iv) Late PTB (34 to 37 Weeks). It is the birth that takes place between 34 and 37 weeks of pregnancy.

Health Impact of PTBPTB is the main risk factor for newborn mortality and morbidity. It is a leading cause of neonatal mortality and morbidity across the world and also the second major cause of child deaths under the age of five years . It arises between 5 and 10% of all deliveries and involves 70% of neonatal mortality and up to 75% of neonatal morbidity. Premature infants are more likely to suffer than normal birth and are at higher

risk of brain paralysis, sensory impairment, respiratory failure, and so on. More than $13 billon of premature cost for maternity service is anticipated only in the USA . Most survivors of PTB face serious problems, often a lifetime of disability, including learning disabilities, visual, and hearing problems. In fact, babies born premature have more health problems compared with babies born at term birth. Term birth refers to babies that are born at 37 to 40 weeks of gestation. Furthermore, babies born at preterm are reported to be at an elevated risk of long-term health problems. Unfortunately, after many years of research in obstetrics, yet the rate of PTB has not decreased. Birth weight is generally associated with PTB and results in its own categorization. Usually, birth weight is simpler to measure precisely and is a first estimation of gestational age. Obviously, the most challenging issue in Gynaecology and Obstetrics is how to control the preterm delivery in pregnant women.Feature Selection (FS)The term feature selection in the machine learning, also known as feature subset selection, refers to the process of selecting a subset of excellent features during construction of the predictive model. The presence of redundant and irrelevant features in any datasets (especially in medical datasets) can reduce the accuracy of the model's prediction and also have the negative impact on the performance of the model. The main goal of any feature selection method is to select the best subset of features by removing redundant and irrelevant features from the datasets in order to reduce the training time and enhance the classifier's predictive performance. In fact, feature selection is typically used as a preprocessing step in data mining. There are three standard approaches of the feature selection algorithm, namely, filter method, wrapper method, and embedded method. For more details about feature selection, one may refer to.(i) Filter Method. The filter method measures the relevance of features based on the nature of data. The selection of features is independent of the classifiers used. The filter method is much faster compared with the wrapper method and provides an average accuracy for all the classifiers used. Some of the examples of filter methods are information gain, chi-square test, variance threshold, and so on. (ii) Wrapper Method. The wrapper method finds the best subset of features based on a specific machine learning algorithm that we are trying to fit on a given dataset. The evaluation criteria are simply the predictive power of the particular classifier. The wrapper method has

higher performance accuracy compared with the filter method but requires more computational time to find best features for a dataset with high-dimensional features. Some of the examples of wrapper methods are forward selection, backward elimination, genetic algorithms, and so on.

Related Works:

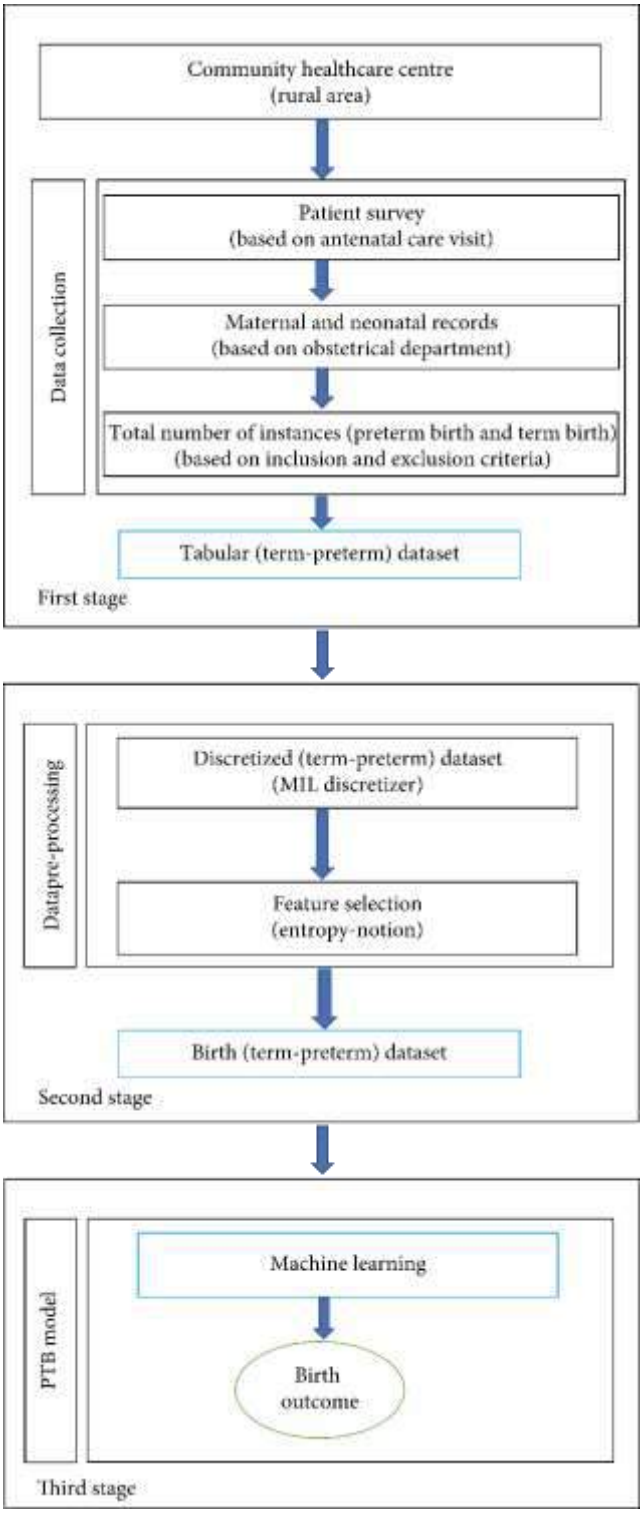
This section focuses mainly on the existing methodologies related to prediction of PTB using machine learning, statistical analysis, and data mining techniques. Some of them are discussed in this section. The study of Mercer al. was designed to develop a risk-score-based model for predicting PTB. The model can be trained using a multivariate logistic regression technique to explore various risk factors using clinical data available between 23 and 24 weeks' gestation. Goodwin et al. employed the machine learning model to generate 520 predictive rules for PTB with the application of data mining techniques. The study in discussed the deep learning models for predicting preterm delivery using existing electronic medical records (EMRs) of mothers available in healthcare centres.

Weber etal. performed a cohort study to predict spontaneous preterm. The prediction of PTB was performed using numerous classifiers, namely, K-nearest neighbours, lasso regression, and random forests. This study has taken into the consideration of demographic, race-ethnicity, and maternal characteristics. Mailath-Pokorny et al. explored the predictive features for preterm delivery that occurs within 2 days after admission and before 224 days of gestation using the multivariate logistic regression model. The predictive features considered are age of the mother, gestational age during admission, maternal history, vaginal bleeding, cervical length, preterm history, and preterm premature rupture of membranes (PPROM) in their study. Son and Miller presented a prediction model for PTB using cervical length measurement in women with a singleton gestation. To accomplish better predictive performance, they attempted to determine the best cut points of cervical length .

Elaveyini et al. explored the major risk factors of preterm birth using artificial neural networks. PTB prediction was based on the feed-forward

backpropagation algorithm. Over the past decades, majority of research studies have been done to enhance the accuracy of prediction of PTB . Researchers are continually making their best efforts to analyse and explore the principal risk factors for preterm delivery . The present article focuses on the machine learning approaches for prediction of birth cases in rural community.

The Proposed Framework: Risk Prediction Conceptual Model (RPCM) Based on novel feature selection (entropy-notion) approach and several studies in , RPCM is carefully designed to predict the risk of PTB in pregnant women. The workflow of the framework consisting of three stages (Stage-I, Stage-II, and Stage-III) is depicted in Figure 2, and then its each component is detailed.


Community healthcare centre
(rural area)
Data collection
Patient survey
(based on antenatal care visit)
Maternal and neonatal records
(based on obstetrical department)
Total number of instances (preterm birth and term birth)
(based on inclusion and exclusion criteria)
Tabular (term-preterm) dataset
First stage
Datapre-processing
Discretized (term-preterm) dataset
(MIL discretizer)
Feature selection
(entropy-notion)
Birth (term-preterm) dataset
Second stage
PTB model
Machine learning
Birth
outcome
Third stage

Conclusion :

In this study, the proposed model (RPCM) can be used for prediction of PTB based on excellent features (text-based symptoms) available in obstetrical data. The work focuses on feature selection (entropy-notion) approach by applying machine learning classifiers (DT, LR, and SVM) in order to classify all birth cases into term birth and PTB. Comparing the performances of the classifiers, it is evident that SVM classifier is the most suitable classifier as it achieves an accuracy of 90.9%. According to the findings of this study, the identified risk factors (excellent features) will be helpful in the prediction of PTB, especially in rural community. The developed model supports the decision-making process in maternity care by identifying and alerting the pregnant women at risk of preterm delivery thereby preventing possible complications, reducing the diagnosis cost, and ultimately minimizing the risk of PTB. The present system can be regarded as a successful innovation in Obstetrics to give clinical support to patients during pregnancy consultations. In particular, RPCM claims to assist healthcare professionals to make effective and timely decisions without consulting specialists directly.

The limitation of the present research is that the risk factors for PTB are limited in size and dataset is small, which could be increased to improve the performance of the PTB prediction in the future studies. However, expert knowledge and clinical judgement may still be needed to interpret this risk and take appropriate action in individual cases.

References:

1. H. Blencowe, S. Cousens, D. Chou et al., "Born Too Soon: the global epidemiology of 15 million preterm births," *Reproductive Health*, vol. 10, no. 1, p. S2, 2013.View at: Publisher Site | Google Scholar
2. L. Liu, S. Oza, D. Hogan et al., "Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis," The Lancet, vol. 385, no. 9966, pp. 430–440, 2015.View at: Publisher Site | Google Scholar
3. B. Govindaswami, P. Jegatheesan, M. Nudelman, and S. R. Narasimhan, "Prevention of prematurity," Clinics in Perinatology, vol. 45, no. 3, pp. 579–595, 2018.View at: Publisher Site | Google Scholar

4. L. J. E. Meertens, P. van Montfort, H. C. J. Scheepers et al., "Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation," Acta obstetricia et gynecologica Scandinavica, vol. 97, no. 8, pp. 907–920, 2018.View at: Publisher Site | Google Scholar
5. M. J. Perez, J. J. Chang, L. A. Temming et al., "Driving factors of preterm birth risk in adolescents," American Journal of Perinatology Reports, vol. 10, no. 3, pp. e247–e252, 2020.View at: Publisher Site | Google Scholar
6. S. Shrestha, S. S. Dangol, M. Shrestha, and R. P. B. Shrestha, "Outcome of preterm babies and associated risk factors in a hospital," Journal of the Nepal Medical Association, vol. 50, no. 180, 2010.View at: Publisher Site | Google Scholar
7. C. Catley, M. Frize, R. C. Walker, and D. C. Petriu, "Predicting preterm birth using artificial neural networks," in In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), pp. 103–108, Dublin, Ireland, June 2005.View at: Publisher Site | Google Scholar
8. A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," BioMed Research International, vol. 2015, Article ID 370194, 16 pages, 2015.View at: Publisher Site | Google Scholar
9. R. Raja, I. Mukherjee, and B. K. Sarkar, "A systematic review of healthcare big data," Scientific Programming, vol. 2020, Article ID 5471849, 15 pages, 2020.View at: Publisher Site | Google Scholar
10. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," IEEE Access, vol. 5, pp. 8869–8879, 2017.View at: Publisher Site | Google Scholar
11. A. Mansoul, B. Atmani, and S. Benbelkacem, "A hybrid decision supportsystem:applicationonhealthcare,"2013, http://arxiv.org/abs/1311.4086View at: Google Scholar
12. B. K. Sarkar, "A two-step knowledge extraction framework for improving disease diagnosis," The Computer Journal, vol. 63, no. 3, pp. 364–382, 2020.View at: Publisher Site | Google Scholar

13. M. Seera and C. P. Lim, "A hybrid intelligent system for medical data classification," Expert Systems with Applications, vol. 41, no. 5, pp. 2239–2249, 2014.View at: Publisher Site | Google Scholar
14. A. Bhardwaj and A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model," Expert Systems with Applications, vol. 42, no. 10, pp. 4611–4620, 2015.View at: Publisher Site | Google Scholar
15. P. J. Lisboa and A. F. G. Taktak, "The use of artificial neural networks in decision support in cancer: a systematic review," Neural Networks, vol. 19, no. 4, pp. 408–415, 2006.View at: Publisher Site | Google Scholar