NEW HORIZON
COLLEGE OF ENGINEERING
Autonomous College, Affiliated to VTU | Approved by AICTE New Delhi & UGC

50 Golden
YEARS OF EXCELLENCE IN EDUCATION

IJSRSET

TECHNO & SCIENCE ACADEMY
The International Open Access Publisher

Third National Conference on Advancements in Computer Science and Engineering

Organised by
Department of Computer Science and Engineering,
New Horizon College of Engineering
In Association with
VTU, Belagavi & Computer Society of India

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Third National Conference on Advancements in Computer Science and Engineering
NCACSE-2021

26th May, 2021

Organised by

Department of Computer Science and Engineering,
New Horizon College of Engineering,
Bellandur Main Rd, Kaverappa Layout,
Bengaluru, Karnataka, India
https://newhorizonindia.edu

In Association with

VTU Belagavi & Computer Society of India

&

International Journal of Scientific Research in Science, Engineering and Technology
Online ISSN : 2394-4099 | Print ISSN : 2395-1990

Volume 9, Issue 3, May-June-2021

Published By

Chief Patron

Dr. Mohan Manghnani
Chairman, NHEI, Bangalore

Patron

Dr. Karisiddappa
Vice Chancellor, VTU, Belagavi

Dr. A S Deshpande
Registrar, VTU, Belagavi

Dr. Manjunatha
Principal, NHCE, Bangalore

Advisory Committee

Mr. Darren Ong
Vice President, APJ, Vertica

Dr. Parameshachari B D
Professor & HoD – TCE, GSSIET

Dr. R Jagadeesh Kannan
Professor & Dean, CSE, VIT, Chennai

Dr. Prakash G
Asst. Professor (Sr.Gr.), ASE, Bangalore.

Dr. Amarjeet Singh
Prof & Dean-Academics, NHCE, Bangalore

Dr. Gopalakrishnan
Professor & Dean-R&D, NHCE, Bangalore

Conference Chair

Dr. B Rajalakshmi
Professor & HOD-CSE, NHCE, Bangalore

Conference Co-Chair

Dr. Pamela Vinitha
Professor, CSE, NHCE, Bangalore

Dr. S Sridevi
Asso. Professor, CSE, NHCE, Bangalore

Ms. Lakshmi S Hanne
Asst. Professor, CSE, NHCE, Bangalore

Organizing Committee

Dr. Clara Kanmani A
Dr. Thirukkumaran R
Dr. Vinodha L
Dr. Rachana P
Dr. Jaya R
Dr. Vaishali M Deshmukh
Dr. S J Subhashini
Dr. Senthil Kumar R
Dr. Anidha Arulanandham
Dr. S Nagendra Prabhu
Dr. A Suresh
Mr. Siva Balan N
Ms. Asha Rani Borah
Ms. Shanmuga Priya S
Ms. Pramilarani K
Ms. Alpha Vijayan
Ms. Soja Rani S
Mr. Santosh Kumar B
Ms. Deepti Rai
Ms. Suganya.R
Mr. Manjunatha Swamy C
Ms. Uma N
Ms. Subhashree Rath
Mr. K Kiran Kumar
Mr. Dinesh G
Ms. Yogitha
Mr. Bhaskar S V
Ms. Tinu N S
Ms. Chempavathy
Ms. Nirali Arora
Ms. Harshitha H

ABOUT NEW HORIZON COLLEGE OF ENGINEERING

New Horizon College of Engineering (NHCE) is an Autonomous college affiliated to Visvesvaraya Technological University (VTU), approved by the All India Council for Technical Education (AICTE) & University Grants Commission (UGC). It is accredited by NAAC with 'A' grade & National Board of Accreditation (NBA). It is one of the top Engineering Colleges in India as per NIRF rankings – 2020. New Horizon college of Engineering is located in the heart of the IT capital of India, Bangalore. The college campus is situated in the IT corridor of Bangalore surrounded by MNCs and IT giants such as Intel, Accenture, Capgemini, ARM, Symphony, Wipro, Nokia, JP Morgan and Cisco to name a few. NHCE has Centre of Excellence with reputed industries like IBM, HPE, VMware, Schneider Electric, SAP, Quest Global, and CISCO.

NHCE has a scenic and serene campus that provides an environment which is conducive for personal and intellectual growth. Students are supported through mentoring and counseling systems. The management offers scholarships to meritorious students. At NHCE, from the moment a student walks into the campus, he/she is well guided to know his/her strengths and choose an area of functional specialization. NHCE has a unique distinction of achieving 100% admissions in all its courses year after year.

ABOUT THE DEPARTMENT

The department of Computer Science and Engineering at NHCE offers graduate, post graduate and PhD programs. It strongly emphasises on learning fundamentals and analysing latest technology to develop student's creativity and encouraging them to take up projects with varying levels of complexity. To meet the fast dynamic technological changes, efficient training sessions are provided on curriculum subjects and beyond. The course work is designed in such a way that it closely intricate with the vision of the department to provide excellent quality education for the student with necessary skills to build computing systems for a "better world".The department ensures continuous skill-enhancement and holistic development of students through various club activities, workshops, conferences, seminars and so on. The department has two industry sponsored Centres of Excellence, HP CoE on Big Data & Data Analytics and IBM open power CoE on Data Science and Machine Learning.

CONFERENCE PREAMBLE

The Third National Conference on Advancements in Computer Science and Engineering (NCACSE-2021) provides a common forum for deliberations, sharing of recent trends, advancements and research in the areas of Computer Science.

This conference will be a very good platform for academia, researchers, industry practitioners and technologists to discuss and present recent advances and research outcomes in the field of Computer Science and Engineering.

CALL FOR PAPERS:

The NCACSE-2021 invites full length original research contributions not submitted to any other journal /conferences. We invite technical papers on, but not limited to, the following domains.

- Image Processing
- Big Data Analytics
- Machine Learning and Deep Learning
- IoT
- Trust, Privacy and Security
- Cloud Computing
- Computational Intelligence
- Data Mining
- Wireless Sensor Networks
- Embedded Systems
- Network Communications
- Gesture Recognition
- Social Network Analysis

SUBMISSION GUIDELINES:

- All papers should be original, unpublished, not submitted to any other journal/conferences, in PDF & DOC format as per the guidelines and IEEE template.
- Please mail your original manuscript mentioning your name, contact details & e-mail id to

About College

New Horizon College of Engineering is an Autonomous college affiliated to Visvesvaraya Technological University (VTU), approved by the All India Council for Technical Education (AICTE) & University Grants Commission(UGC). It is accredited by NAAC with 'A' grade & National Board of Accreditation (NBA). It is one of the top Engineering Colleges in India as per NIRF rankings, ARIIA – 2020 and an ISO 9001:2008 certified Institution. New Horizon college of Engineering is located in the heart of the IT capital of India, Bangalore. The college campus is situated in the IT corridor of Bangalore surrounded by MNCs and IT giants such as Intel, Accenture, Capegemini, ARM, Symphony, Wipro, Nokia, JP Morgan and Cisco to name a few.

NHCE has a scenic and serene campus that provides an environment which is conducive for personal and intellectual growth. The infrastructure acts as a facilitator for the effective delivery of the curriculum. NHCE boasts of state -of-the -art facilities for its students. The students are given utmost encouragement in their areas of interest by providing hi-tech facilities backed by faculty support. The institute places highest priority on innovative programs of instructions that include both traditional class room theory and professional skills training. There is a strong impetus on overall personality development of the students with emphasis on soft skills. Students are supported through mentoring and counseling systems. The management offers scholarships to meritorious students. At NHCE, from the moment a student walks into the campus, he/she is well guided to know his/her strengths and choose an area of functional specialization. This enables students to concentrate their efforts and energies to gain the competitive edge. NHCE has a unique distinction of achieving 100% admissions in all its courses year after year.

NHCE has Centre of Excellence with reputed industries like Adobe, HPE, Vmware, Schneider Electric, SAP, Quest Global, CISCO.

CONTENTS

A Review on Design and Development of IoT Based Pulse Oximeter

Raksha T Murthy[1], Sharadhi S[1], Varshini R[1], Rashmi S[1], Ambika V[1]

[1]Department of Computer Science and Technology, Vidyavardhaka College of Engineering Mysuru, Karnataka, India

ABSTRACT

Measuring the heart rate, oxygen saturation level and blood pressure plays a vital role in human body. In many critical conditions measuring these parameters is very important when a patient is in critical condition so pulse oximeter is really necessary. Early pulse oximeter could only measure saturation level of oxygen. Later pulse oximeter evolved to measure pulse rate along with heart rate with temperature inbuilt control. Red and infrared sensors are used in pulse oximeter to detect heart beat and saturation level of the oxygen. In this paper pulse and saturation of oxygen can be calculated using different types of methods that are easy and accurate to calculate. Alert system will alert the patient when there is drop of any level of Spo2 and heart rate.

Keywords- Iot, Spo2, Heartrate, Blood Pressure, Pulse rate.

I. INTRODUCTION

[2] Pulse oximeter is a medical device that is usually clipped in fingertip used to measure the heart rate and saturation level of oxygen in blood. Pulse oximeter uses red and infrared sensors. The light emitted by these sensors is absorbed through skin and later they get absorbed through blood. Then amount of light absorbed by our oxygenated and deoxygenated hemoglobin present in blood is calculated in the machine to detect the saturation of oxygen and pulse rate of a person. There will be OLED display which is used to display the result of pulse rate and oxygen saturation level calculated. Earlier pulse oximeter used red and green filters to find the oxygen saturation level which was not accurate. Later the original pulse oximeter was developed which calculated accurate value of oxygen saturation. Today pulse oximeter available will calculate both the heart rate and saturation level of oxygen that are very accurate. According to change in environment it will also affect the working function of body like increase in heart rate so temperature is inbuilt in this device to accurate the working efficiency of the device.

Blood pressure can be added to available device to make the device even accurate and helpful. If there is any increase above the normal value in any of the parameters like heart rate pulse rate alert will be activated. Using message alert system and GPS we can detect condition and location of patient. Alert is sent to the relatives of the patients and in case of any emergencies message is sent to nearby hospital. OLED display is used to see the output of the devices.

II. LITERATURE SURVEY

[1] Nowadays one of the cause for death is due to heart related disease, so the cardiac values are need to be continuously monitored. Heart rate and SpO2 count are measured using sensor. Heart rate and oxygen saturation levels are detected using pulse oximeter with LEDs and photo detectors. Measured

data will be digitalized using converters and then it is stored in cloud. If there is any variation in cardiac value an alert is sent automatically. Patient's location will be tracked using GPS.

[2] This paper explains the design of the portable, wearable, real time health monitoring system using IOT. Heart rate, ECG Signal, body temperature and SpO2 of the patients are monitored using sensors. Main feature of this system is to send measured parameters to doctor's smart phone. MAX30102 heart rate sensor, signal-lead ECG sensor AD8232, LM35 analog temperature sensor are connected to Node MCU and Arduino UNO microcontrollers. After collecting data this will be sent to blynk server. This mobile application provides an affordable health monitoring system for the patients.

[3] This paper explains health monitoring sensor system which is used to measure SpO2 count, heart rate and body temperature. Led driver circuit is used to enhance the accuracy of pulse oximeter. This wearable wireless sensor is affordable. This paper deals with developing a system which has accurate pulse oximeter sensor, temperature sensor, heart rate sensor integrated with Arduino to measure SpO2 , pulse rate and body temperature. Data collected from this device is transmitted to internet or mobile at low cost. Results of this paper are compared with commercially available ChoiceMMed pulse oximeter and we can observe maximum deviation of 2% which can be accepted.

[4] This paper describes the implementation of wearable, high mobility SpO2 and heart rate monitoring system. This device consists of heart rate and SpO2 sensors which are connected to Node MCU microcontroller which provides gateway to the internet. Data measured using these sensors will be stored in database. User needs to register in website, which displays heart rate and oxygen saturation (SpO2) results. These data can be observed in both

patient and doctor side and this is used to measure the progress of the patient and also used to provide right treatment. In this system it is found that there is a difference of +-2.8bpm for heart rate and +-1.5% for spo2.

[5] This paper says about IOT based powerline technology. This implementation uses Iot based powerline communication. This system uses star network topology. In this system all the slave system are interconnected to master system through a network. The heart of this pulse oximetry is spectrophotometry. The red and infrared sensor are used in which these rays penetrate through our skin and reaches our vain in which they will take count of oxygenated and deoxygenated haemoglobin.
The Master Control System is simply a collection of graphical development environment, which recovers all the information given by the slave systems. This software is organized in the form of account where it shows all information of pulse oximeter along with time and graphical representation of data (acquired from PSS), information regarding to people being examined. All the date will be stored in database server. The data can be accessed through internet using web server. The communication rules uses handshaking property. For each request, the clients responds with an acknowledgement. The powerline slave System receives and dispatches the data through this network. The data cannot be accessible to doctors are patients. Security in this system is good.

[6] This paper tells about home based tele-health care services for measuring saturated oxygen and heart rate based on the client-server architecture. Client receives the oxygen count and heart rate through pulse oximeter[hardware components] which is stored in the server side which can be accessible to all doctors where they will have their own account encrypted with password. They can access this information through Wi-Fi communication. Configuration for users are done using Bluetooth.

This project is especially designed for monitoring the hypoxia patients those who suffers from lower amount oxygen level inside the blood. If there is any variation in the oxygen it can be monitored by the specified family doctor through the server side of internet. Whenever there is emergencies the doctor may be not available which may lead to complications to patients. The error showing in counting the saturation level of oxygen in range from 100% to 73% is less than 1 percent. When range is about 73 to 50% is 2 percent.

[7] During the hospital nature, patient record habitually formed with costly resources .The intension of the project is to provide both hardware and program solution in user friendly manner, the easiest solution is when the instruments automatically makes the measurements and stores the results into the cloud and informs the patients at final level or pre warning level. The pulse oximetry system carry-out statistical calculations based on the Beer-Lambert Law to calculate the ratio of blood that is saturated with oxygen.

[8] This paper describes the implementation of oxygen saturation (spo2) and heart rate monitoring system using Internet of Thing (IOT).In IOT system it enables preventive care and encourage automation to reduce the risk of human mistakes. These project is mainly concentrated on the safety of women and child preventive healthcare.
As a part of implementation of the system there is a portable device that serves to retrieve psychological data of mother and child from several sensor and send them to the database, some of the psychological data are heart rate, SPO2 ,fetal heart rate, ECG, photoplethysmogram (PPG).In these paper they were implemented heart rate and SPO2 level.

[9] The economy of china has been disturbed due to COVID-19 so healthcare has gained more importance. So Health Monitoring System using IOT has a

solution for it. The main aim of this project is to implement patient health monitoring system that tracks patient health issues and takes the help of internet to inform their family and friends. The microcontroller are attached to the sensor to track the position of the patient. These values which are tracked by the microcontroller are been displayed to LCD screen. When certain epidemic is spread it is impossible to reach the doctor in particular time. So to avoid this, if health monitoring system is given to patients they can be monitored from their own places. This system monitors heart rate, humidity and temperature. These values is sent via wireless communication which are also displayed on LCD.

[10] This paper tells about how to measure the heart rate and absorbance of oxygen using wireless communication unit. Here heart rate is detected using peak detection algorithm and the correlation function method. In this method pulse rate can be calculated by measuring threshold value when it reaches its peak value it obtains the ac cycle maximum value and when it reaches zero then the new threshold value is calculated. Between the pulses that has been calculated the counter will acquire their period and finally pulse rate can be calculated. By using the value they can also calculate saturation of oxygen level. Once these values are calculated the result is sent to the wireless microcontroller. This system requires microcontroller to be connected to wireless network and mobiles since input is not present in microcontroller. Through Wi-Fi connection are established to phone and the microcontroller and they can communicate it easily. The obtained result will show difference less than 2%.

[11] This paper tells about the wearable device that consists of sensors, circuits and microcontroller. This is very light weight and easy to wear. Wearable device is also very comfortable to wear while doing the exercise. This wearable device also consist of memory chip, MCU, ADC etc. Here device is

connected to phone which can be monitored through our phone. Mobile displays all the parameters of heart rate and oxygen level. While doing exercise it keep tracks of all cycle of exercise, if there is any increase above the normal value the server will detect the drop rate and the alert message will be sent to phone. This systems accuracy rate is about 98.96%.

[12] This paper is designed to calculate the temperature, oxygen absorbance and pulse rate of patient. Here temperature sensors and pulse sensors are used and they are connected to arduino Uno. The values that are detected here are stored in cloud server and they are analyzed. The values of the heartbeat, oxygen absorbance in blood and the temperature can be seen through the led display. If there is any drop in any other values cannot be monitored here like blood pressure and the pulse rate cannot be detected here. Error detecting while calculating spo2 is 0.89% while in heart rate 3.906%.

[13] This paper mainly talks about parameters like temperature , ECG and heart rate in which these are connected to arduino Uno in which they monitor the values. This paper also uses amplifier circuit to increase the strength of the signals as they are all weak. When device is connected to internet it will be connected to cloud server in which the data that are calculated are sent through. Here cloud server and the arduino communicate with each other and send and receives the data and are saved in cloud platform. This system is very cost effective which will not be easily available. This paper does not tell us about calculation of blood pressure and it does not have any alert system.

[14] Small changes in some variable in physical body such as oxygen congestion, temperature, BP and heart rate results in disease. The differences in these parameters are need to be measured to determine whether the certain disease is present or not. Using IOT and wireless sensors patient's data can be

delivered to the doctor. This plays a major role in helping doctors to take necessary action to send medical help. The patient's data is processed and stored in cloud. Several benefits in storing data in cloud area availability, reliability and convenience. This paper provides secure solution for health records that are stored in cloud. Sensors are used to measure heart rate, spo2 and body temperature. Microcontroller is used for filtering, encode and decryption of health variables to the cloud. This paper AES algorithm which is used to secure patient data before storing it in cloud. Node MCU microcontroller is used for connectivity over Wi-Fi to cloud. This system can also send alert mails to patient's relatives and friends.

[15] In this paper the IOT is used to provide smooth medical facilities to the patients and facilitates doctors and hospital as well. This system would be confidential to the family members and their doctors about the current status of patients and medical records. This paper they used raspberry pi in which they acts as a controller to demonstrate a remote health monitoring system. The information of a patient's status is saved to the cloud and it will be showed to the authorized user through online. It will also send the alert message to the patient's family members and also doctors at the time of emergency.

III. CONCLUSION

IOT along with some sensors is used to implement health monitoring system which is able to measure heart rate, SpO2 and body temperature. The paper tells us about powerline communication used and tele health care services used to implement heart rate and spo2 count. These device only calculates the heart rate and oxygen level saturation and does not consider other parameter. Blood pressure sensor can be added to detect blood pressure in blood and also alert system can be used to send alert message to patient's family. This system can implement heart

rate , saturation level of oxygen and also sends an automatically send patients health information alert along with patient's location to the doctor. So this system helps the patients to get a right treatment at the earliest stage.

IV. REFERENCES

[1]. Dhanurdhar Murali, Deepthi R Rao, Swathi R Rao and Prof Ananda M –" Pulse Oximetry and iot based cardiac monitoring integrated alert system.

[2]. Mustafa A Al-Shiekh and Ibrahim A Ameen-" Design of mobile healthcare monitoring system using Iot technology and cloud computing"

[3]. Mian Mujtaba Ali, Shyqyri Haxha, Munna M. Alam, Chike Nwibor and Mohamed Sakel – "Design of internet of things(IOT) and android based low cost health monitoring embedded system wearable sensor for measuring spo2, heart rate and body temperature simultaneously"

[4]. R. R. Adiputra, S. Hadiyoso and Y. Sun Hariyani – "Internet of things: low cost and wearable spo2 device for health monitoring"

[5]. Giovanni Bucci, Fabrizio Ciancetta, Edoardo Fiorucci, Andrea Fioravanti and Alberto Prudenzi – "A Pulse Oximetry Iot system based on Powerline Technology"

[6]. Thai M. Do, Nam P. Nguyen and Vo Van Toi – "Development of a Pulse Oximeter for E-health Applications"

[7]. P. Szakacs-Simon, S.A. Moraru and L. Perniu – "Pulse Oximeter Based Monitoring System for People at risk"

[8]. T M Kadarina and R Priambodo – "Monitoring heart rate and spo2 using things board IOT platform for mother and child preventive healthcare"

[9]. Prajoona Valsalan, Tariq Ahmed Barham Baomar, Ali Hussain Omar Baabod - "IOT based health monitoring system"

[10]. Adan Torralba Ayance, Hector Santiago Ramirez, Jose Migue Rocha Perez and Carlos Gerardo Trevino Palacios – " Wireless Heart rate and oxygen saturation monitor"

[11]. Eka Adi Prasetyo Joko Prawiro, Chun-l Yeh, NaiKuan Chou, Ming-Wie Lee and Yuan- Hsiang Lin - "Integrated wearable system for monitoring heart rate and step during Physical activity"

[12]. M T Tamam, A J Taufiq and A Kusumawati - "Design a system of measuring of heart rate, oxygen saturation in blood and body temperature with non-invasive method"

[13]. Shivleela Patil, Dr. SanjayPardeshi - "Health monitoring system using IoT"

[14]. Ali I.Siam, Atef Abou Elazam, Nirmeen A El-Bahnaswy, Ghada El Banby, Fathi E. Abd El-Samie -"Smart health monitoring system based on IOT and cloud computing"

[15]. Shubham Banka, Isha Madan and S.S.Saranya - " Smart healthcare Monitoring using IOT"

Pneumonia Detection Using Deep Learning

Dr. Pamela Vinitha Eric[1], Abhishek Mesta[1], Mithun S[1], Jhanavi D[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Pneumonia is a form of acute respiratory tract infection (ARTI) that affects the lungs which is caused by bacteria, viruses or fungi. Pneumonia is the leading disease that occurs more in children of age below 5 years. Every year almost 7,00,000 children are victimised for this disease. Hence the accurate diagnosis of such a disease is of high importance. So, the expert radiologists role is crucial to identify the disease through chest x-ray images. But, in certain situations the doctors fail or there are no expert radiologists available in developing countries. There is a requirement of a software-based support system to detect Pneumonia using Chest X-ray images to provide early diagnosis for the infected person. So, the aim of this project is to develop a software system to detect the disease Pneumonia using Chest x-ray images. This is achieved by using multiple convolutional neural network layers where the chest x-ray images are tested, trained and validated. The inception v3 model is a CNN which is 48 layers deep and is used to extract the high-level features from the images. The test result obtained showed that the software is classifying the infected and non-infected images.

Keywords— InceptionV3 model, Convolutional Neural Network, Deep Learning, Machine Learning, Data Mining

I. INTRODUCTION

Pneumonia is one of the diseases that has been noted throughout human history. The bacteria that caused pneumonia was discovered in 1881 which is known as streptococcus pneumonia. It is an infection that affects lungs which produces chills, fever, coughing and difficulty in breathing in an individual. Generally, this disease is found in children of age below 5 years. Infants are affected very soon since the immune system is weak. Each year, around 1.5 million children die because of pneumonia. It is mostly seen in developing nations. Pneumonia can be cured in early stages using antibiotics. Expert radiologists can

identify if an individual is normal or infected. But in certain cases radiologists may fail to identify through naked eyes or sometimes there might be scarce doctors in under developed or developing nations which can lead to death. Therefore, there is a need for a computer-based support system to identify the infected and non-infected. Hence, the project aims to achieve this by proposing a model using Deep Learning method. By using InceptionV3 which is a Convolutional Neural Network. In the world of computer vision Inception V3 helps in the classification of objects. It was also used in life sciences, where it aids in the research of Leukemia. The dataset for the project is collected from Kaggle

which contains images of both infected and normal lungs. Dataset is divided into 3 categories such as train, test and validate. The training, testing and validating of datasets happens in layers of convolutional neural networks.

II. RELATED WORK

An approach for the diagnosis of Pneumonia from chest X-ray images using deep learning

In 2019, Enes Ayan proposed a methodology for the analysis of pneumonia from chest x-beam pictures utilizing profound learning strategy. In this paper the analysis of Pneumonia was drawn nearer with two notable CNN models which are Xception and VGG16. They have utilized adjusting and move leaming strategies for their preparation stage. For the exchange learning strategy they have used DenseNet-121 layer convolutional neural organization it is otherwise called CheXNet. The organization of theirs is prepared with 10.000 front facing view chest X-beam pictures which had 14 distinct infections and accomplished 90% of accuracy. Further, the interaction has been partitioned into two organizations, their first organization depends on the Xception model and the subsequent one is Vgg16 based model. At that point, the two models are analysed dependent on their presentation and the outcomes shows the Xception model takes a lead in the exhibition when contrasted with the Vgg16 model in diagnosing pneumonia. The test aftereffects of their methodology in Vgg16 organization and Xception network prevailing at the exactness with 82% and 87% separately.

An approach to detect Pneumonia clouds by using image processing methods

In 2017, Sharma proposed a way to deal with distinguish Pneumonia mists by utilizing picture preparing techniques. In which they proposed the

strategy to analyse the sickness from clinical pictures. They dealt with 40 simple Chest X-beam which is parcelled to typical and pneumonia. They created native calculations to edit and recognize the limit locale of the lung from the chest x-beams. At first, every one of the 40 pictures were resized to ideal size for computational purposes. Further, histogram leveling was performed on something very similar to improve the picture contrast. Afterward, the mid region was edited by utilizing native calculations. At that point, the locale of the lung which isn't overcast was recognized by utilizing the Otsu picture thresholding strategy and the territory was processed further. Then, the non-shady district proportion to that of the removed absolute region locale of the lung, would give the cloud arrangement data part in the lung. In this way, they arrived at the exactness of around 76% to analyse pneumonia.

An automatic detection of Pneumonia analysing ultrasound digital images

In 2016, Barrientos proposed a programmed recognition of pneumonia investigating ultrasound computerized pictures. This methodology depends on the examination of examples present in rectangular fragments from ultrasound advanced pictures. They gathered the conclusion of pneumonia which is 15 and this was from the 23 lung ultrasound pictures in which 8 were Further, the comparative vectors were separated from casings and ordered those equivalents as the past one for both positive and negative. Then, required locales were recognized inside pictures. At last, from the districts of interest highlights were extricated and are given to the enn calculation utilized. This neural organization calculation utilizes sigmoid enactment work which has three layers i.e., input covered up and yield. Finally, to accomplish the Pneumonia recognition, preparing and testing measures were performed. Hence, this paper presumes that the affectability and particularity of the undertaking came to 91.52% and 100%.

Rethinking the inception architecture for the computer vision

In 2015, Christain Szegedy with the reference to the paper Rethinking the origin engineering for the PC vision. We discover the creators zeroing in on the higher performing of convolution neural organizations, for example, VGGNet and Google Net. It is seen that, however Inception design of Google Net cycle to performs well considerably under exacting memory imperatives and computational budget, when contrasted and VGGNet and AlexNet, which requires high computational expenses have concerns in regards to its compositional intricacy. This paper gives not many general standards and thoughts of improvement, which is useful in increasing convolution neural organizations effectively and proposes Inception V3 model, which moderately has unobtrusive calculation cost contrasted with less complex and more solid design. A methodology for the article disclosure which is capable limit the calculation time was cultivated with high accuracy. The methodology was regular to build up a defy revelation system which is around speedier than any past approach. This paper unites present day calculations. portrayals, and pieces of information which are nonexclusive and have more extensive application PC vision and picture planning. Finally, this paper presents a lot of tests on an inconvenient face identification dataset which has been extensively considered.

III. METHODOLOGY

Here is the detailed working and assessment steps of our venture. Our venture depends on the sickness pneumonia which is identified utilizing Chest X-beams first and foremost have gathered the dataset (i.e., chest x-beam pictures) from kaggle. This Project as the setup of Keras which is an open source Deep Neural Network system with the TensorFlow backend which is utilized to assemble and prepare the Convolutional Neural Network then we have utilized jar programming interface structure for the mix of frontend and backend, flask system is likewise coordinated for unit testing by utilizing google collaborator for running the python code, it gives quicker TPU and RAM size of 12gb for the quicker preparing measure.

The dataset of our undertaking comprises of three principal envelopes which are preparing, trying, and approval. These envelopes further contain the two subfolders Pneumonia (P) and Normal (N) Chest X-beam pictures. In this model a sum of 5,856 X-beam pictures are utilized. The chest X-beam was overseen from the patient's clinical routine consideration. The extent of information ought to be adjusted for that interaction it is been allocated to the preparation and approval set and the first information will be modified. Dataset is been partitioned into training, testing and approval set.Here.by taking the complete number of pictures i.e. 4,898 that is been given to the preparation set and 3,567 pictures is been given to the approval set to expand the precision of the undertaking.

In pre-processing of information stage, while for the information expansion i.e., Fig-1, The cycle utilized is in Image Data generator class which is utilized for the resizing of the chest x-beam pictures. In a more specialized manner this class is utilized for the even interpretation which chips away at the width moving by 0.1% and for the upward interpretation which deals with the stature moving by 0.1% itself which is equivalent to the width shift and the class mode utilized is all out. In extension, a shear stretch out of 0.2 percent of the image focuses are in a counter clockwise heading. The zoom broadens subjectively zoom's the photos at the extent of 0.2 percent, and after that finally, the photos were flipped equally.
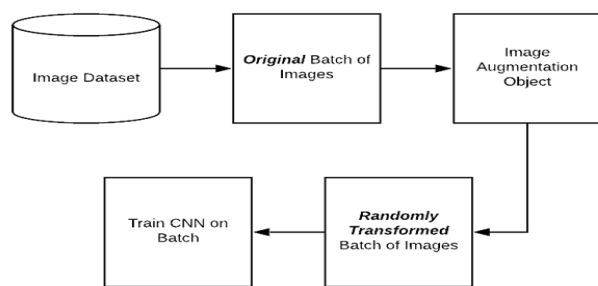
Fig. 1 Data augmentation I Keras

Further, the component extraction is likewise done in the information pre-processing stage. For the extraction of highlights is finished by applying the InceptionV3 model which is a Convolutional Neural Network that is 48 layers deep. Inception v3 TPU preparing hurries to coordinate with precision which is delivered by GPU working of comparative design. The model has been effectively prepared on v2-8, v2-128, and v2-512 setups. The model has accomplished more noteworthy than 78.1% precision in around 170 ages on every one of these cycles and it has been prepared by the pretrained model of organization. This pretrained model contains more than a large number of pictures from the ImageNet information base which is now prepared, and that organization can characterize pictures into 1000 items or more. Therefore, this Inceptionv3 model has taken in a plentiful component extraction for an example scope of pictures. The organization has contribution of a picture size of 299-by-299. The functioning model of the initiation v3 is gotten from Fig-2.
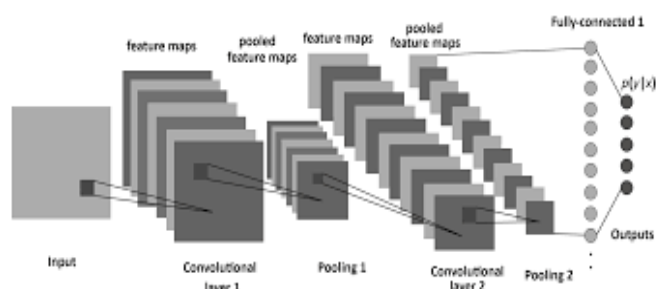
Fig. 2 Architecture of InceptionV3 model.

In this process, it comprises of a few layers and those layers are max pooling, fully associated layer keras conv2D and actuation work.

Here, the capacity of max pooling is utilized for the dimensionality decrease. The principal objective of this is proportional down the given information portrayal. A 2D Convolution Layer as Keras Conv2D under it, this layer makes a convolution piece that is gasping for air with numerous layers of information which delivers a tensor of yields. Every one of these conv2D and max pooling capacities are applied on various occasions to prepare the given dataset for more productivity and exactness.

When all convolution and pooling layer capacities on the information are done, at that point the information is prepared to the completely associated layer. In this layer, it extricates every one of the contributions of one layer and interfaces it to the each enactment unit of the following layer, this is utilized to give the last yield. For the order cycle we have a SoftMax classifier. It utilizes the cross-entropy misfortune. The evacuation of crude class scores into standardized positive qualities that whole to one is finished by SoftMax classifier. So further the cross-entropy misfortune can be applied. For the assemblage of the code, we have utilized Adams analyser calculation to give an enhancement which can deal with meagre inclinations on boisterous issues.

IV. CONCLUSION AND FUTURE WORK

Early and effective conclusion is fundamental to lessen the passing brought about by pneumonia. In the spaces where radiologists are in lesser numbers, our framework to identify pneumonia will get itself valuable. This apparatus can be of huge assistance to a considerable lot of the oppressed all throughout the planet. Further upgrades to our work which could be performed are:

1. To discover an approach to investigate pneumonia through the ultrasound video of the lungs.
2. To build up an application where not just the medical clinics gain admittance to it, even the

typical individuals can download the application and check for the actual outcomes.

3. To plan the calculation to break down rgb x-beam. By and by it is intended to work for grayscale pictures.

4. To expand the exactness of the undertaking.

V. ACKNOWLEDGEMENT

VI. REFERENCES

[1]. Enes Ayan, Halil Murat Unver, "Diagnosis of Pneumonia from Chest X-Ray Images using Deep Learning", 2019.

[2]. Abhishek Sharma, Daniel Raju, Sutapa Ranjan, "Detection of Pneumonia clouds in Chest X-ray using Image processing approach". 2017.

[3]. Ronald Barrientos. Avid Roman-Gonzalez, Franklin Barrientos, Leonardo Solis, Malena Correa, Monica Pajuelo, Cynthia Anticona. Roberto Lavarello, Benjamin Castaneda, Richard Oberhelman, William Checkley, Robert H. Gilman, Mirko Zimic, "Automatic detection of pneumonia analyzing ultrasound digital images".2016.

[4]. Vanhoucke Sergey Architecture for Shlens, "Rethinking the Inception Vision",2015.

[5]. Inception V3(2020): CNN. Available from: https://en.wikipedia.org/wiki/Inceptionv3

[6]. https://www.kaggle.com/paultimothymooney/chest-xray-pne umona/version/2

Eye Control of Devices for Paralysed and Semi Paralysed

Shruthi M[1], Sushmitha[1], S Selva Lavanya[1], Manjunatha Swamy C[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Eye control of devices for paralysed and semi paralysed is a novel idea to aid differently abled and paralysed persons to use a device like a computer easily. It is common knowledge that most conventional input devices like the mouse, keyboard, etc cannot be easily used by a paralysed or a semi paralysed person. This project aims to track where the user is gazing on the screen and use this as input to operate the device. The system initially looks out for a face in the video that is fed. The system then detects the iris in the eye and puts its position in a coordinate system. This is mapped onto the screen. Blinks are used in place of clicks. The duration of blink to click is set higher than for a natural blink of the eye. Once the system is set up, we aim for it to continuously track the eye gaze and take that as input from the user. It can help the users communicate their needs, make distress calls and operate a machine.

Keywords— Eye Detection, Human Computer Interaction, Computer Vision, Face detection, Text to Speech

I. INTRODUCTION

Several people all throughout the planet are experiencing an assortment of actual incapacities that keep them from carrying on with a typical life. It might bring about numerous problems, for example, amyotrophic parallel sclerosis (ALS), cerebral loss of motion, horrendous cerebrum injury or stroke that may bring about the deficiency of muscle developments in the body, making the individual incapacitated. As such, these individuals are in a vegetative state with just their eyes moving. Such individuals are totally subject to others for essential capacities like taking care to change in position.

Impaired individuals can't move anything aside from their eyes. For these individuals eye development and flickers were found to be the most effective means of communicating with the outside world through PC.

This exploration points in building up a framework that can help the actually tested by permitting them to connect with a PC framework utilizing just their eyes. Human-computer interaction is becoming more prevalent in our daily lives. The user's eye movement can be used as an input source that is easy, natural, and high-bandwidth.

The implementation explained in the project is idiosyncratic because unlike pre-existing represented implementations there are no usages of hardwares like semiconducting cathodes, or any other medium of luminescence source to detect the eyes. The only source of hardware required in this project is a working PC or laptop with a working webcam, which makes it heuristic and viable. The implementation is first carried out by taking pictures of the user with the help of the webcam which processes the input image individually and compares

the shift of the iris. After the image is processed thoroughly the estimation of the iris position is carried out and the tool maps the cursor position onto the screen.

As just a direct result, we'll be ready to discern where the user is gazing depending on the positioning of the mouse cursor. The tool will produce the corresponding events by focusing the gaze at a position for a few seconds. Users then will pick and click the appropriate functions.

The eye tracking market is estimated to increase at a CAGR of 24.5 percent during the forecast period, from USD 368 million in 2020 to USD 1,098 million by 2025.

One of the primary driving drivers for the eye tracking market is the rising demand for eye trackers in the healthcare vertical, particularly for assistive communication applications.

Figure 1 shows the bar graph of World Eye Tracking market size by product.
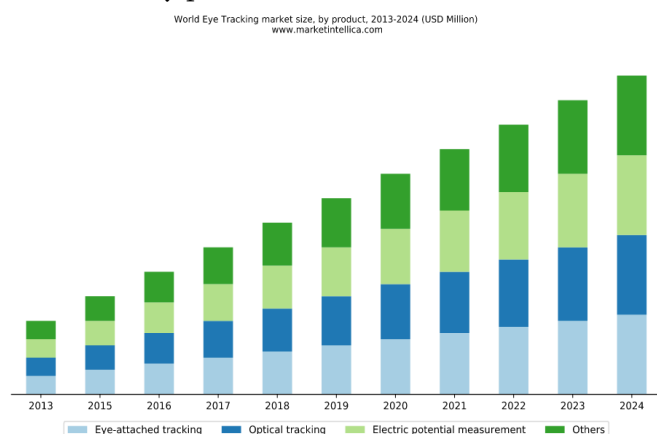
Fig. 1 World Eye Tracking market size

II. PROPOSED SYSTEM

Most disabled people find it hard to use conventional computer systems. We intend to create a software that uses a webcam camera to detect the user's eye gaze and find out where he/she is looking at the screen.

The software first identifies the user's face. This is done using the dlib library functions namely get_frontal_face_detector and shape_predictor. The system uses landmark points provided by the dlib library to find the user's face. Once the system finds a valid face it moves on to eye detection. If a valid face is not found the system intimates the user. Once the face is identified the next step is to identify the eyes. Eye detection is done using the dlib library functions similar to face detection using the landmark points. Once the face and eye is detected, the next step is to detect the iris. This is done using the Opencv library's findContours function. Once the iris is identified, the eye is divided into 4 quadrants. After this we find the iris location in the eye. The relative position of the gaze on the screen is mapped by finding out which quadrant the iris is at a given time. We calculate the position on the screen by passing the coordinates of the iris to a function and map the position on the screen. The user will be able to perform the click operation with a blink.

Advantages
* IR LED is not used; therefore, no harm is caused to the eye of the user.
* Expensive hardware is not required hence, the system becomes more feasible.
* Installation is fairly simple because of the lack of excess hardware.
* System is not mounted on the head of the user or on glasses. This reduces discomfort.
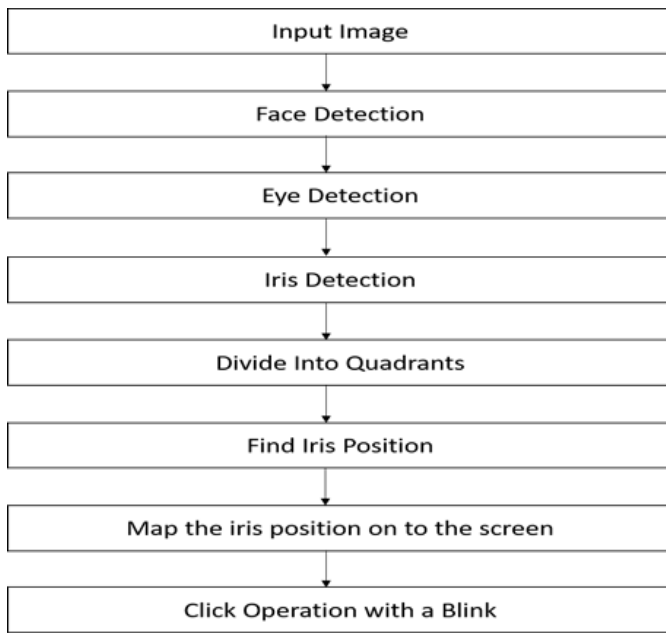
Fig. 2 Outline of Proposed System

Figure 2 displays the working of the proposed system. Main components in our system are face recognition, eye recognition, iris detection and mapping the iris position on the screen.

A. Face and Eye Recognition

Face Detection is done using dlib library. Dlib is a library of machine learning algorithms that has been designed to be highly modular, easy to execute, and simple to use. The dlib uses a pre-trained landmark facial detector that traces the facial points on a user's face using sixty eight coordinates landmark points.

We use OpenCV to capture images. The captured frame is then converted from RGB format to grayscale format. A function is used to find faces in the image which was converted to grayscale image. If no faces are found then the system intimates the user. If faces are found then for each face we find the landmarks and convert it into a numpy array. We find eyes using the landmark points and convert it into an array.

B. Iris Detection

Once the eyes and face are detected we go on to detect the iris. Using NumPy, we build a new black mask whose size is similar to our webcam photo.

Using cv2.fillConvexPoly we trace the (x, y) coordinate points of the left and right eyes from the keypoint array shape on the coverup we built. This function takes points as numpy array, colour and image as parameters and gives back an image in which the section between the points will be coloured with the specified colour. After we do this we will be left with an eye area in white colour. Using the OpenCV functions we were able to segment out the eyes. In the segmented fragment all the (0,0,0) pixels are converted to (255,255,255) so that the dark part left will be the eyeball. OpenCV's findContours function is used to find the iris and its center.

C. Mapping Iris position

As we were able to successfully detect face, eye and iris position, we were able to extract the iris coordinates with respect to the eye. Using these coordinates we can calculate the position on the screen. When the eye is closed for a few seconds we register it as a click at that coordinate. This value is set higher than average blinking time to distinguish between a natural blink and a blink that is intended to be registered as a click.

III. TECHNICAL ASPECTS

A. Python

Python is an open source, interpreted, object oriented language. Features such as dynamic typing and binding make it suitable for rapid application development. Apart from that python is easy to learn since it has almost English like commands. Python is preferred for this task since it has a vast collection of libraries and is easily compatible with libraries written in other languages. It is one of the most preferred languages for computer vision.

B. Open CV

OpenCV is short from for open source computer vision. It is a cross platform library used to perform image and video processing. It has various inbuilt

functions for face recognition, object recognition, image transformation, etc. In this project we are using OpenCV 2 in Python. Our main use of OpenCV is to capture real time video input and find the iris using the contour detection function cv.findContours().

C. Dlib

Dlib is an open source library principally based in C++. Here we are using it with python for facial recognition. Dlib uses 68 point landmarks of the face to perform face detection. It has high accuracy and will enable us to check if the eye or mouth is open or closed easily. The library has a pre-trained model which uses the 68 coordinate (x,y) as a template.
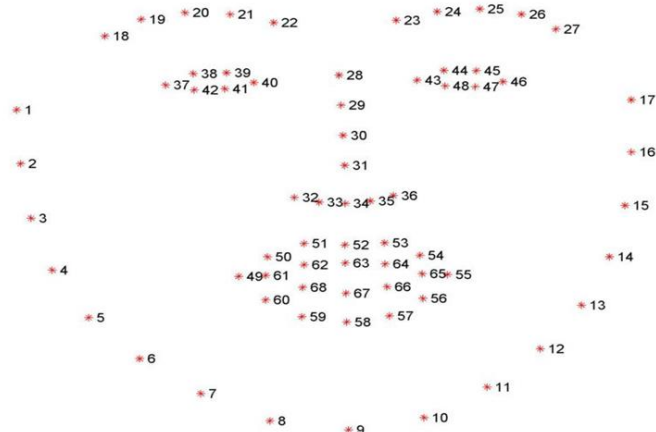
Fig. 3 Dlib 68 point coordinate system

We use Dlib to initially check for the face and eye visibility. Dlib is also used to isolate the image of the eye from the rest of the face and find the iris inside it. It also makes it easier for us to detect blinks.

D. Tkinter

Tkinter is Python's widely accepted standard GUI (Graphical User Interface) package. We are using it to create an UI interface for the user to interact with. It provides a powerful object oriented approach to build windows and place widgets like buttons, text views, text boxes, etc inside it.

E. Text to Speech

The pyttsx3 library is used to convert text to speech. It is compatible with Python 2 and 3 and it uses the pyttsx3.init() function to invoke a reference to the pyttsx3.Engine instance. It is a simple way to convert the typed text to audio. We are utilising this feature to let the user communicate with a caretaker with ease.

IV. RESULTS

Figure 4.1 shows the initial face and eye recognition check before the user interface launches.
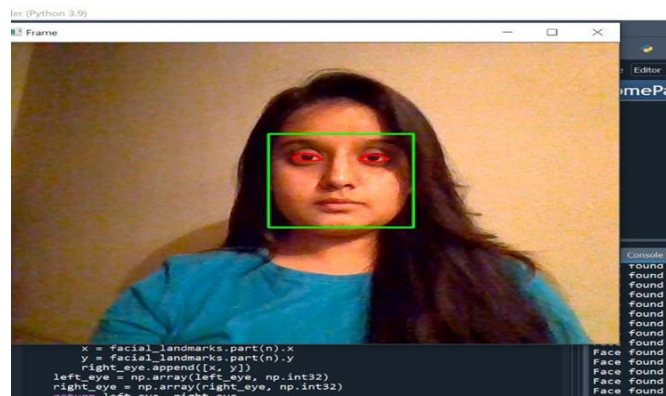
Fig. 4.1 Face Detection

Figure 4.2 depicts the iris detection and thresholding of the image. The mouse pointer is moved according to the iris position.
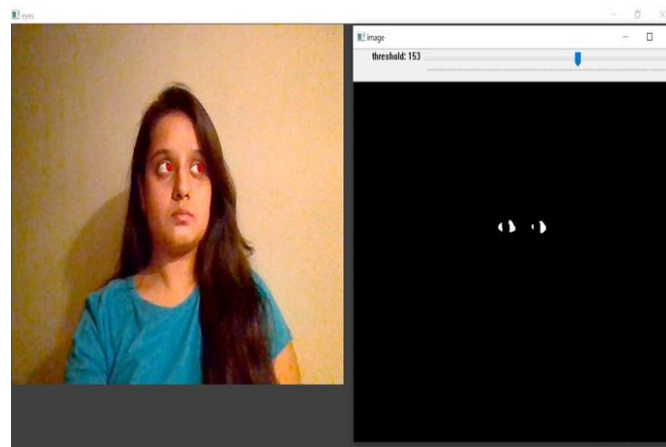
Fig. 4.2 Iris Detection

V. CONCLUSION

The dependence on conventional input devices like the mouse and keyboard is still prominent. This acts as a huge disadvantage to people who are differently abled. Most available solutions require expensive

hardware like high configuration cameras or IR lights that shine into the eye to know the position of the pupil. These techniques often require a head mounted sensor which might cause discomfort. We aim for this project to reduce the complexity of hardware setup and use inbuilt computer cameras and lighter applications to identify the gaze location. Apart from only helping the differently abled, this technology has wide untapped potential in psychology, advertising, marketing, industrial application, safety control (drowsiness detection in drivers), etc.

Benefits:-
● Doesn't require expensive hardware.
● Easier to set up and use.

Disadvantages:-
● Needs good lighting for accuracy.
● Some system specific values must be fine-tuned for each system for optimal performance.

VI. ACKNOWLEDGEMENT

VII. REFERENCES

[1]. Singh, B., Kandru, N., & Chandra," Application control using Eye Motion", 2014 . The IEEE website. [Online]. Available: http://www.ieee.org/

[2]. Salunkhe, P., & Patil, A. R, "A device controlled using eye movement ", 2016 . The IEEE website. [Online]. Available: http://www.ieee.org/

[3]. Mohamed Nasor, 1Mujeeb Rahman K K, Maryam Mohamed Zubair, Haya Ansari, Farida Mohamed, ""Eye-Controlled Mouse Cursor for Physically Disabled Individual", 2018,The IEEE website. [Online]. Available: http://www.ieee.org/

[4]. Vinay S Vasisht, Swaroop Joshi, Shashidhar, Shreedhar, C Gururaj, "Human computer interaction based eye controlled mouse" 2019, The IEEE website. [Online]. Available: http://www.ieee.org/

[5]. Dr. Menua Gevorgyan, "OpenCV 4 with Python Blueprints: Build creative computer vision projects with the latest version of OpenCV 4 and Python 3, 2nd Edition", Dr. Arsen Mamikonyan.

[6]. Joseph Howse, "OpenCV: Computer Vision Projects with Python: Computer Vision Projects with Python: Develop computer vision applications with OpenCV ",Prateek Josh.i

[7]. The Dlib library documentation website.[Online]. Available:http://dlib.net/python/index.html

[8]. The Tkinter documentation website.[Online]. Available:https://docs.python.org/3/library/tk.html

[9]. The OpenCV documentation website.[Online]. Available:https://docs.opencv.org/master/d5/d98/tutorial_mat_operations.html

[10]. Text to speech converter.[Online] Available:https://www.geeksforgeeks.org/python-text-to-speech-by-using-pyttsx3/

QUBIT – Smart Education

Srinivas R[1], Harikrishnan GS[1], Saravana Kumar B[1], Ms. Soja Rani S[2]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Sr. Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Using online learning mode makes things easier for the students as well as teachers, as it helps plan the training activities calendar which the user can share with your learners, teachers, and co-administrators. By doing this the user can maintain and improve the learning process, which is a time consuming process when done manually. Online learning is education that takes place over the Internet. It is often referred to as "e-Learning" or "Online Classroom" based learning. They are generally conducted through a learning management system, in which students can view their course syllabus and academic progress, as well as communicate with fellow students and their course instructor. Online education enables you to study or teach from anywhere in the world. This means there's no need to commute from one place to another, or follow a rigid schedule. On top of that, not only do you save time, but you also save money, which can be spent on other priorities. The virtual classroom is also available anywhere there's an internet connection.

There's often access to very diverse material such as videos, photos, and eBooks online as well, and tutors can also integrate other formats like forums or discussions to improve their lessons. And this extra content is available at any moment from anywhere, which will offer you a more dynamic and tailor-made education. All lectures and needed materials are provided via online platforms, so you'll easily access them from the comfort of your home. Through online classes, each and every student can gain an education. When the students are not able to go outside due to some serious reasons, then they can easily study from the online classes. Online education enables the teacher and the student to set their own learning pace, and there's the added flexibility of setting a schedule that fits everyone's agenda. Studying online teaches the student vital time management skills, which makes finding a good study-life balance easier. Online education has also been the principal form of distance education but now, it is also changing the instructions on higher education as it is now becoming a global agent in higher education. Advancements in technology learning have contributed to the enhancements of generations of face-to-face learning and generations of distance education.

Keywords — Online Education, Teacher, Student, Flexibility, Advantage, Learning, Virtual Classroom, Save Money, Save Time, Focussed Learning.

I. INTRODUCTION

Online learning is education that takes place over the Internet. It is often referred to as "e-Learning" or "Online Classroom" based learning. They are generally conducted through a learning management system, in which students can view their course syllabus and academic progress, as well as communicate with fellow students and their course instructor. It is a platform where a lecturer can conduct classes similar to traditional classroom sessions, through online platform and also conduct assessments and assignments to evaluate the understanding and the performance of the students.

It enables efficient and more focused learning than online courses, as there is a live instruction and support from the faculty in order to clear doubts and assist the students in their learning process. Online Classes are typically a mix of video recordings or live lectures supplemented with readings and assessments that students can complete on their own time.

COVID-19 has forced a sudden migration to online learning with little time to prepare for it. As the pandemic accelerated, colleges shifted into emergency mode, shutting down campuses in an effort to prevent the spread of the disease caused by the novel coronavirus and moving academic life online.

Online students typically log in to a learning management system, or LMS, a virtual portal where they can view the syllabus and grades; contact professors, classmates and support services; access course materials; and monitor their progress on lessons.

II. OBJECTIVE

To develop a user friendly and efficient solution for the effective delivery of online education and provide a smooth and seamless online class experience to the teachers and students, and also develop an efficient system to monitor the attendance of students as well as proctor their assessments effectively by using

existing hardware devices and increase the productivity of online learning.

III. SCOPE OF THE PROJECT

The scope of this project is to build a robust and efficient system which enables the effective delivery of online learning to students through a more focused and progress oriented approach. This system can enable a better interaction between the students and teachers and also make the learning process more comfortable for both, and also make it easy to track the process more accurately.
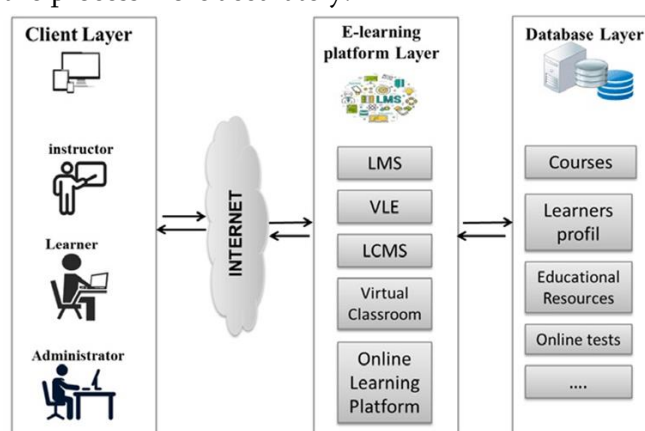
IV. BENEFITS OF ONLINE LEARNING

- Minimized Travel expenditure.
- Minimized Infrastructure costs.
- Better Focus on the screen rather than classroom.
- Reduced disturbances for students from peers.
- Better interaction with the faculty through a single button.

V. ONLINE AND DISTANCE LEARNING

Distance education, also called distance learning, is the type of education of students who may not always be physically present at a school. Traditionally, this usually involved correspondence courses wherein the student corresponded with the school via post. Today, it involves online education. A distance learning program can be completely distance learning, or a

combination of distance learning and traditional classroom instruction.

Many colleges and schools have started online education for international students and even for the local students for their courses. It is very beneficial for people who are willing to pursue master's courses while they are working. The purpose of online education is to provide flexibility, wide range of courses, and more cost efficiency.

Nowadays universities and schools prefer online classes during unfortunate situations like climate and weather changes or pandemic. Since the spread of COVID-19 all schools and universities have started taking their classes online. They have started making everything online including exams, quizzes, assignments and projects. So online education is very helpful for both universities and schools and we can surely say that Online learning is the future of education.

VI. EXISTING SYSTEMS

IV.A.1 GOOGLE CLASSROOM

It is a web service developed by Google for schools that aims to simplify creating, distributing, and grading assignments. The primary purpose of Google Classroom is to streamline the process of sharing files between teachers and students. Students can be invited to join a class through a private code, or automatically imported from a school domain. Google Classroom integrates various features into a single platform to manage student and teacher communication. Teachers can create, distribute and mark assignments easily and they can monitor the progress for each student and can return work along with comments.

Disadvantages

- The main drawback of this scheme is that the solution does not have any feature of proctoring to monitor the performance of a student.

- No features having live sessions for interaction with the faculty.
- No feature to have a visual lecture for clarification.
- No feature to conduct exams and quizzes and proctor the students for the same.
- No platform to enforce the regular attendance of students.
- Cannot create meeting for conducting classes
- No options for viewing student's marks and attendance
- The teacher cannot share their screen and explain concepts

IV.A.2 FACIAL RECOGNITION ATTENDANCE

To maintain the attendance record with day to day activities is a challenging task. The conventional method of calling name of each student is time consuming and there is always a chance of proxy attendance. This system is based.

As the time for corresponding subject arrives the system automatically starts taking snaps and then apply face detection and recognition technique to the given image and the recognize students are marked as present and their attendance update with corresponding time and subject identifier. Use of facial recognition techniques for the purpose of student attendance further this process can be used in exam related uses.

While gathering Information of students it also generates dataset of student faces and stored in folder. In real time scenario the camera is placed near the classroom door which captures live data. Continuous observation improved the performance for estimation of the attendance. 2 cameras increase the accuracy and estimates the seating accuracy as well. Different angles provide a better perspective of recognising a person.

Disadvantages:

- The solution needs a physical camera and needs to be installed at a right position.

- It is suitable only for a physical classroom where real people are in front of the camera, and not in a virtual setup.
- Does not have any other feature other than capturing attendance such as classes.
- The solution external cameras which are expensive and cannot work with laptop cameras.
- No features of assignments and discussions.
- No feature to conduct live sessions and meetings between the teachers and students.

VII. PROPOSED SYSTEM

It is difficult for teachers to have different software which does different tasks like marking attendance, scheduling meeting, sending notes and assignment. The students and teachers need to use multiple solutions simultaneously in order to achieve the complete objective of an online class experience and this leads to lack of coordination and hence mismatch in the pace of the teaching and understanding of the student. Thus we need a complete all-in-one solution for online education. This provides features like online assignment submission, scheduling meetings, maintaining and managing marks of individual students, attendance marking system and proctored exams all in a single platform. This system allows the faculties to share notes to the respective class and sections they handle and allows the students to submit their homework and assignments in their portal. This project uses machine learning for recognizing students through their webcam for capturing and marking their attendance and it is also used for proctoring during exams and quizzes. This system lets the students to view their marks and attendance in their profile and also lets the faculty to scheduling meetings to a particular class and share the invite with the students in the same platforms and hence the students can join the class with the click of a single button.

VIII. REQUIREMENTS

HARDWARE
- Processor: i3, i5, i7
- RAM: 8 GB
- Hard Disk: 500 GB
- Input device: Standard Keyboard and Mouse
- Output device: High Resolution Monitor

SOFTWARE
- Operating system: Windows 10, Linux, MacOS
- Front End: HTML, CSS
- IDE: Visual Studio Code, Android Studio
- Data Base: Firebase/MongoDB/SQL
- Server: Firebase/Apache

IX. MACHINE LEARNING AND DEEP LEARNING

Machine learning is an artificial intelligence (AI) application that gives systems the ability to learn and improve from experience automatically without being programmed explicitly. Machine learning focuses on the creation of computer programs that can access and use data to learn on their own.

Deep learning is a function of artificial intelligence (AI) that mimics the functioning of the human brain in data processing and the creation of patterns for decision-making use. Deep learning is an artificial intelligence subset of machine learning that has networks known as deep neural learning or deep neural networks. A field of artificial intelligence that trains computers to interpret and understand the visual world is computer vision. Machines can accurately identify and classify objects and respond to vision using digital images from cameras and videos and deep learning models.

Recognition of the face is a method of identifying or verifying an individual's identity using their face. Face recognition systems can be used to identify people in photos, video, or in real-time. A facial recognition system is a technology capable of matching a human face against a database of faces

from a digital image or a video frame. Using Python libraries such as Tensor Flow, OpenCV and Dlib, face recognition can be implemented.

X. DESIGN GOALS

Online education is a learning process with the combination of content that is both delivered digitally and through face-to-face learning. Online education contributes to the shifts from traditional face-to-face learning to the use of web technological tools which enhances collaborative learning and presents an entirely new learning platform for students. Online education has also been the principal form of distance education but now, it is also changing the instructions on higher education as it is now becoming a global agent in higher education. Advancements in technology learning have contributed to the enhancements of generations of face-to-face learning and generations of distance education. As to it, when online education develops, it has begun to use different approaches to address diverse goals.

The Major Goals of online education

There are certain goals when it comes to online education and some of these are to:

- Enhance the quality of learning and teaching
- Meet the learning style or needs of students
- Improve the efficiency and effectiveness
- Improve user-accessibility and time flexibility to engage learners in the learning process
- Make the interface more user-friendly to both teachers and students

Online education is vast and an expanding platform with huge prospective in higher education. Since there are many challenges in making online education effective, it is important to know how to manage it and access to the resources.

The Design Goals of QUBIT involve various factors such as:

- Creating a User Interface that is accessible and User-Friendly to both the teachers and the students, as all teachers and students may not be technologically sound.
- Creating a Solution which can be accessed from different platforms and is not constrained by hardware limitations.
- Enabling the efficient learning of the students and effective teaching of the teachers by providing a good channel of interaction between both.
- Using a centralized database that can serve the needs of all the functionalities of the application and across various platforms.
- Providing Secure authentication and maintaining the privacy of the details of the students and teachers.
- Creating an accurate and efficient ML based Image Recognition System that can fulfill the needs of Proctoring as well as capturing the attendance automatically.
- Creating a system that can reduce the overall costs of Learning such as Travel, Books and Materials etc.
- Creating a system that the User can use readily without any need of a tutorial.
- Building a system that is both comfortable and convenient to the Students and Teachers.
- Providing data security and restricting access to those who are not authorized
- Build a system that reduces the manual effort and energy of the students and teachers along with saving their time

FLOWCHART OF THE SYSTEM
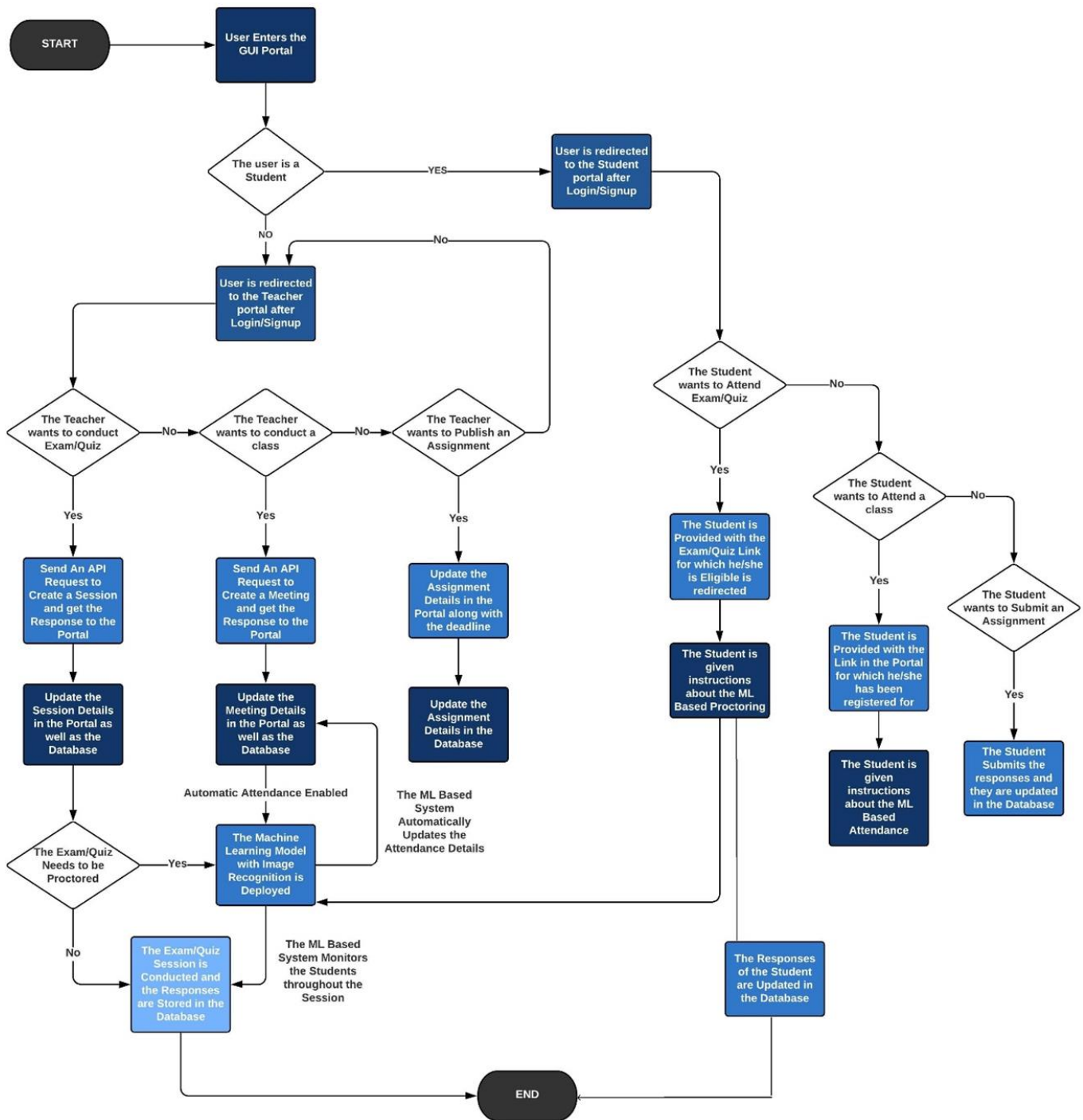
USE CASES IN THE SYSTEM
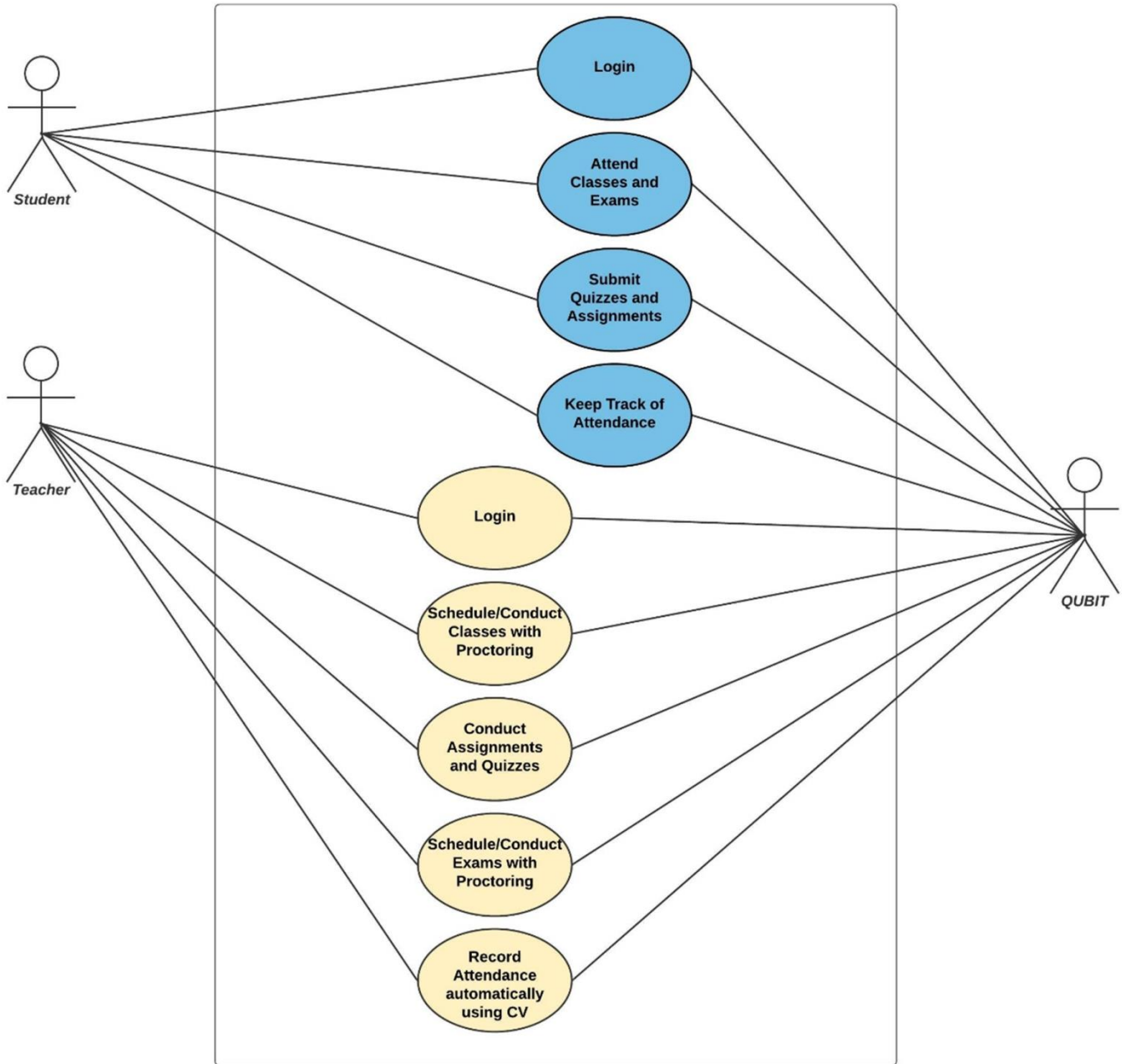
SEQUENCES OF PROCESSES IN THE SYSTEM
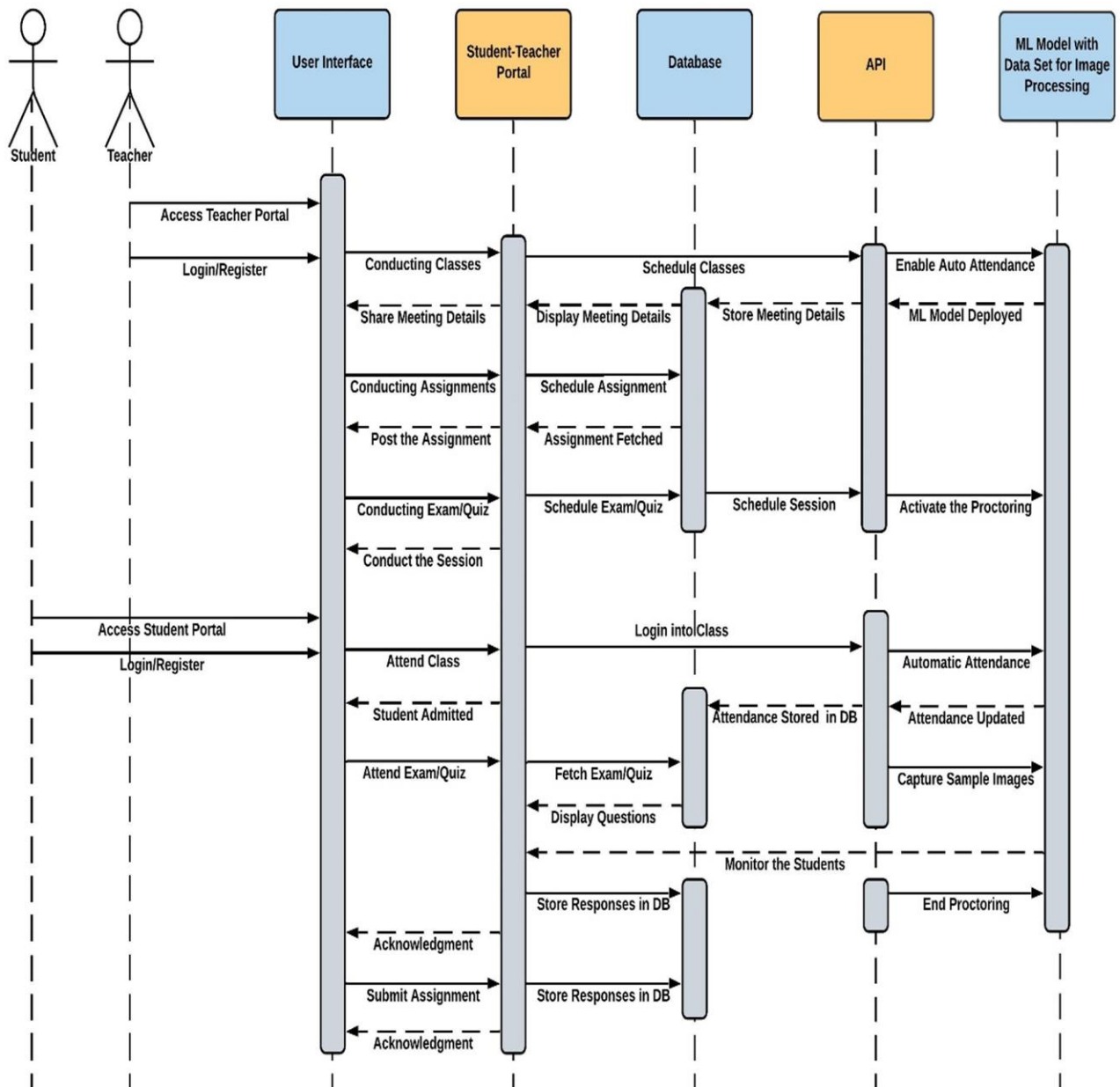
SYSTEM ARCHITECTURE
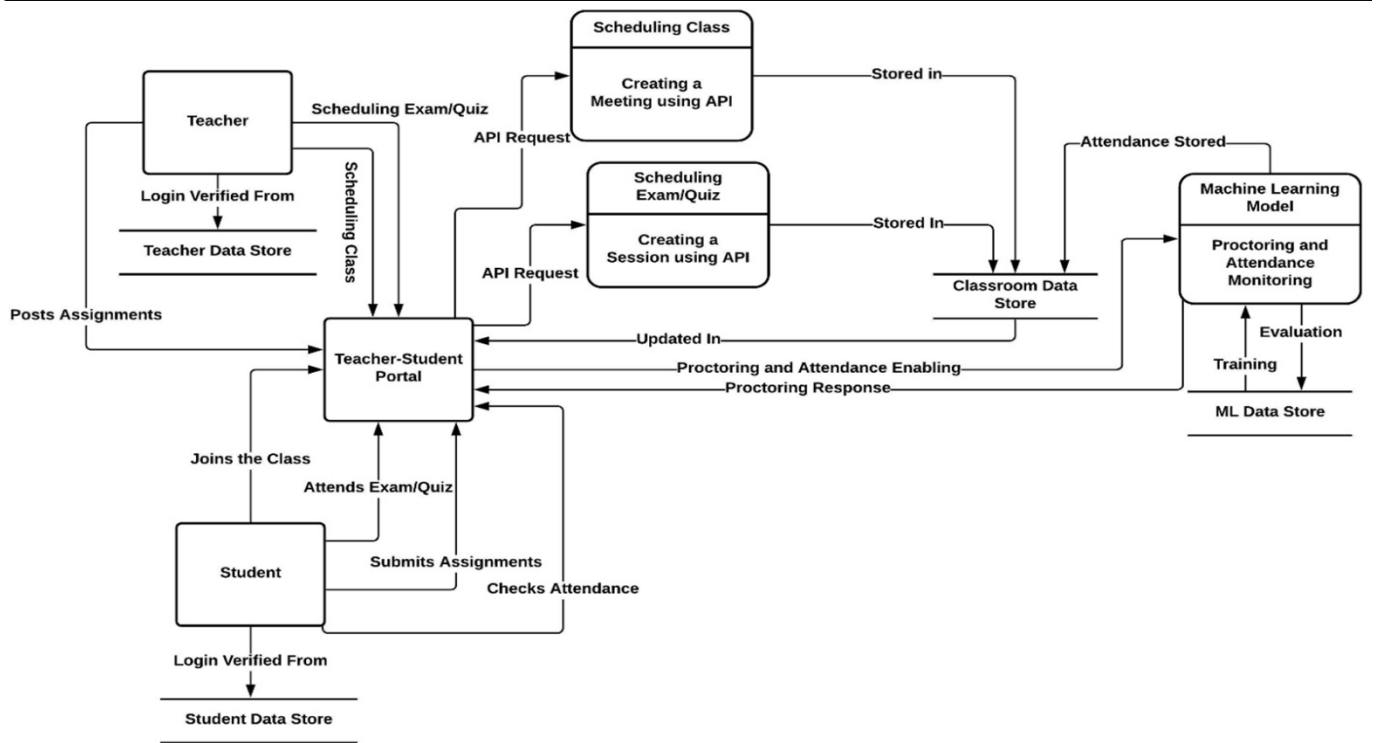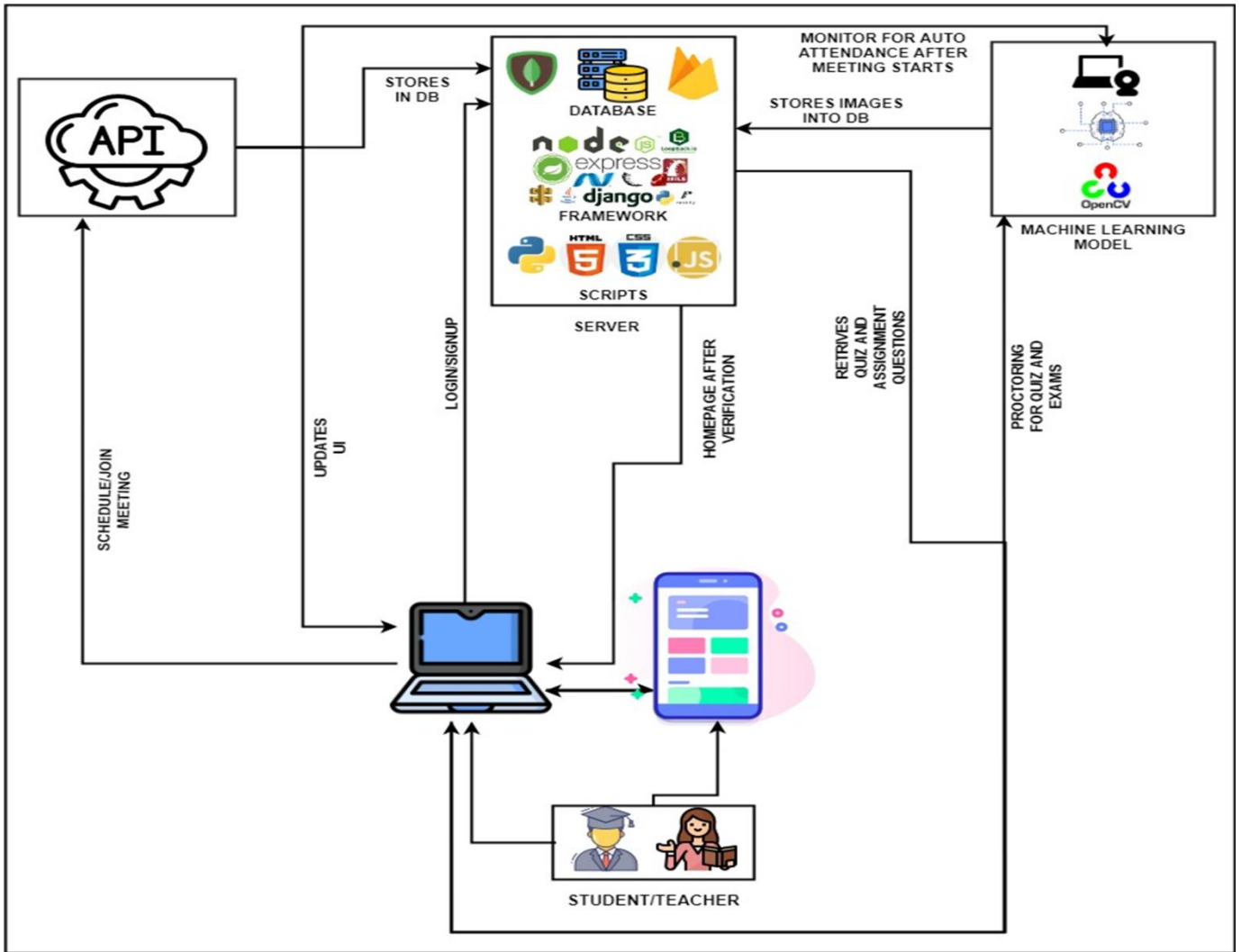
XI. DJANGO

Django is a web development framework that makes it easier to create and maintain high-quality web applications. Django makes the creation process simple and time-saving by eliminating repetitive tasks. Model-View-Template (MVT) is a variant of Model-View-Controller (MVC). The key difference between the two patterns is that Django handles the Controller (software code that governs the interactions between the Model and the View), leaving us with the template. The template is a combination of HTML and Django Template Language (DTL). The below diagram illustrates how each of the components of the MVT pattern interacts with each other to serve request to the user –
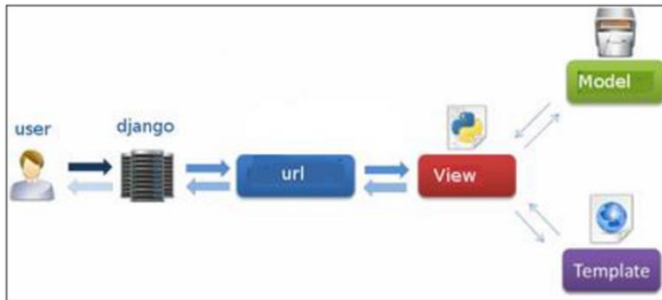
The developer provides the Model, the view and the template then just maps it to a URL and Django does the magic to serve it to the user.

XII. DATABASE

A database is a logically organized collection of data that is stored and retrieved by a computer system. When dealing with more complex databases, systematic design and modelling techniques are often used. The database management system (DBMS) is a software that interacts with end users, programs, and the database itself to collect and analyze data. The database management system (DBMS) software also comes with the required tools for database management. The database, the database management system, and the associated applications make up a "information system."

XII.A.1 SQLITE (DEVELOPMENT)

SQLite is a C library that provides a framework for managing relational databases (RDBMS). Unlike many other database management systems, SQLite is not a client–server database engine. Rather, it becomes part of the final product. SQLite is ACID-compliant and, like PostgreSQL, implements the bulk of the SQL specification. SQLite uses a dynamically typed SQL syntax that does not guarantee domain integrity.

XII.A.2 POSTGRESQL (DEPLOYMENT)

PostgreSQL is an efficient, open source object-relational database framework that uses and expands the SQL language, as well as a number of features that enable it to securely store and scale even the most complex data workloads. PostgreSQL was created in 1986 as part of the POSTGRES project at the University of California at Berkeley, and the main platform has been actively developed for more than 30 years.

PostgreSQL has a strong reputation for its dependable design, data confidentiality, robust feature set, extensibility, and the open source community's commitment to providing performant and innovative solutions on a continuous basis. Since 2001, PostgreSQL has been an open-source database management system that runs on all major operating systems, has powerful add-ons like the common PostGIS geospatial database extender, and is ACID-compliant. It's no surprise that PostgreSQL is the open source relational database of choice for many individuals and businesses.

PostgreSQL has a range of features that can help developers create applications, administrators protect data privacy and create fault-tolerant environments, and you manage data, regardless of the size of the dataset. In addition to being free and open source, PostgreSQL is extremely extensible. Many of the features of the SQL standard are supported, but some have slightly different syntax or functionality. More progress toward conformity can be predicted over time. As of September 2020, PostgreSQL version 13

conforms to at least 170 of the 179 required features for SQL:2016 Core conformance. As of this date, no relational database complies with this specification fully.

They Are Data Types, Data Integrity, Concurrency, Performance, Reliability, Disaster Recovery, Security, Extensibility, Internationalization, Text Search

XII.A.3 MONGODB

MongoDB is a free, open-source, cross-platform document-oriented database. MongoDB is a NoSQL database that uses JSON-like documents and optional schemas to store data. MongoDB is a database developed by MongoDB Inc. and released under the terms of the Server Side Public License (SSPL). MongoDB allows us to search by field, range, or normal expression.

Queries can return unique fields from documents as well as JavaScript functions that the user specifies. Queries may also be configured to return a random sample of responses with a fixed number of responses. Fields in a MongoDB book can be indexed using primary and secondary indices. MongoDB provides high availability by replica sets. A duplicate package is made up of two or more versions of the data. Any member of the replica-set will take on the role of primary or secondary replica at any time. Both writes and reads are done on the primary replica by default. Secondary replicas can hold a copy of the primary's results thanks to built-in replication. When a primary replica fails, an automated election is held to determine the secondary should take its place as the primary. If desired, read operations can be served by secondaries, but data is only compatible by default. If the repeated MongoDB implementation only has one secondary member, a separate daemon called an arbiter must be added to the array. It only has one task: to determine the outcome of the upcoming primary election. As a consequence, an idealised distributed MongoDB implementation needs at least three separate servers, even though only one main and one secondary server is used. MongoDB can scale horizontally due to sharding. Individuals choose a shard key, which determines how data in a collection is transmitted. The information is separated into ranges and dispersed through several shards (based on the shard key). (A shard is a master who controls one or more replicas.) The shard key can also be hashed to correspond to a shard, ensuring that data is distributed evenly. MongoDB may be spread through many machines, balancing the load or duplicating data in the event of hardware failure. GridFS, a MongoDB-based file system with data replication and load balancing, can be used to store files across several computers.

XIII.   HEROKU

Heroku is a platform-as-a-service (PaaS) that runs on the cloud and supports a number of programming languages. Heroku, one of the first cloud services, was established in June 2007 and initially supported only Ruby. Java, Node.js, Scala, Clojure, Python, PHP, and Go are also supported. As a result, Heroku is known as a polyglot platform because it enables developers to build, run, and scale applications in a consistent manner across multiple languages. Customers' apps run in Heroku's virtual containers, which provide a secure runtime environment. By Heroku, these containers are referred to as "Dynos." The languages Node, Ruby, PHP, Go, Scala, Python, Java, and Clojure are all supported by these Dynos. Custom buildpacks are also available on Heroku, allowing developers to deploy applications in any language they want. The developer of the app can easily scale it by changing the type of dyno it runs on or increasing the number of dynos it runs on using Heroku. Heroku Postgres is a cloud storage (DBaaS) provider for Heroku that uses PostgreSQL. Forks, supporters, and dataclips are all available in Heroku Postgres, as are continuous security, rollback, and high availability. Heroku Redis is a Redis branded by Heroku with the aim of improving the developer experience. Heroku takes care of everything and runs it like a script.

Customers can use a CLI to handle instances, connect data with Postgres to gain business insights using SQL tools, and see the results.

XIV. METHODOLOGY OF FACE RECOGNITION

The following steps are involved in the process of Face Detection and Recognition:

- **Detecting the face:** Using the built-in camera on a laptop or smartphone and enabling the Face Detection feature in the pre-existing camera using the required application and settings.
- **Identification of reference points:** In the image, the app will detect specific features. The main identification point is traditionally eyes, but developers usually introduce up to 70 points to their solutions. We have to run manual tests and map out these spots so that these actions can be performed automatically by the software later.
- **Frontal face modeling:** The application should be able to remodel the picture to get the frontal image if the face was pictured from non-direct angles. This is done by combining in a single portrait the reference points from original images.
- **Calculating the Descriptor:** The next step is to introduce the face vector as a combination of points, lengths, and quantitative characteristics: the sort of characteristics that mathematically describe the face of a person. It also defines such key factors as hairstyle, gender, age, race, etc.
- **Comparing faces:** The next step is to cross-match the detected faces with others in the database by applying Euclidean distance. The system will create a unique ID for this new face if no matches have been found. If the image matches an existing input, the system will display an ID corresponding to the vector that is present in the database.
- **Facial Recognition:** For an existing face, once the verification from the database is complete, the model should perform the required action. The model must register the new face for a fresh face and perform an action based on that.

XV. TECHNIQUES FOR FACE RECOGNITION

- **Eigenfaces**: It is a Principal Component Analysis (PCA) based classification method to decrease the dimensionality of each image and project its attributes on the new dimensions considered. Finally, by comparing the Euclidean distances between the data obtained for each picture, the final classification is obtained. This method offers reasonably satisfactory outcomes and has a low computational load.
- **Fisherfaces:** The method of Fisherfaces is based on LDA (Linear Discriminant Analysis) and uses data between members of the same class to create a set of feature vectors in which the differences between the different faces (or classes) are increased while the differences are minimized within the same class.
- **Hidden Markov Model (HMM):** A Hidden Markov Model is a statistical model that assumes that the system model is an unknown parameter process of Markov and can be regarded as a simple Bayesian dynamic network. This model is based on the division of regions with facial images (7 nodes) and the probability of a state transition. The algorithm provides a probability that with certain determined characteristics, a given region of the face will follow another.
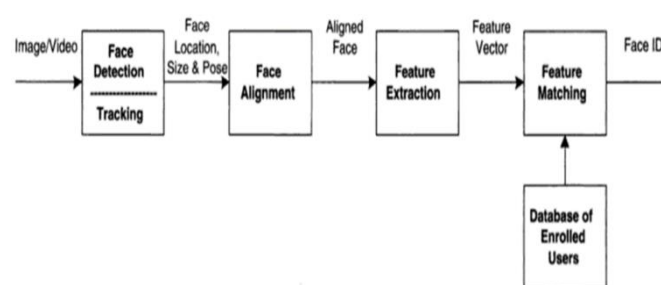
XVI.   FUTURE WORK

The Future Scope of expansion of this project includes the following:

- Technologies such as Augmented Reality and Virtual Reality can be integrated to provide a full-fledged laboratory experience at home, but however that needs the investment of extra devices.
- The Machine Learning model can be extended further to not just detecting the particular person, but also to detect their emotions and determine if they are able to follow the subject while the teacher is conducting the class, and if not alert the teacher in between automatically.
- We can also expand this to observe the attentiveness of the student, and the teacher can keep track if the student is really listening to the class with concentration.
- Payment Gateways can be integrated with the application to enable the students to pay their academic fees as well.
- The system can also have an integrated Coding and Testing environment for students to type their code, compile and run them.
- The system can be expanded for teachers to post tutorial videos along with online live sessions so that the students can refer them later for doubts.
- The application can also have gamified learning process, by which all subjects can be integrated with a game, such as quizzes and other games so that the students are more interested and are excited to learn.

XVII.   CONCLUSION

Online education allows the teacher and the student to set their own pace of learning, and there is the added flexibility of setting a schedule that fits the agenda of everyone. Studying online teaches the student vital time management abilities, which makes it easier to find a good balance between study and life. Online education allows you to study from anywhere in the world or to teach. This means that there is no need to commute or follow a rigid schedule from one place to another. In addition to that, you not only save time, but also save money, which can be spent on other priorities.

The virtual classroom is also accessible wherever an internet connection exists. Each and every student can get an education through online classes. If for some serious reasons, the students cannot go outside, then they can easily study from the online classes. In such cases, online classes help to remove barriers and borders, and students, even sitting at home, can readily gain knowledge. Online classes provide students with high-quality training at their own location. For students to study effectively, online learning is a great solution. There is also access to very diverse content such as online videos, pictures, and eBooks, and tutors can also integrate other formats such as forums or discussions to enhance their lessons. And this additional content is accessible from anywhere at any time, which will give you a more dynamic and tailored education. Online platforms provide all the lectures and necessary materials, so you can easily access them from the comfort of your home.

XVIII.   REFERENCES

[1]. https://conductscience.com/basic-tools-and-techniques-of-data-science/
[2]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm
[3]. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
[4]. "Face Recognition-based Lecture Attendance System", Yohei KAWAGUCHI, Tetsuo SHOJI, Weijane LIN, Koh KAKUSHO, Michihiko MINOH
[5]. "The Application of Google Classroom as a Tool for Teaching and Learning", Izwan Nizal Mohd

Shaharanee, Jastini Mohd Jamil, and Sarah Syamimi Mohamad Rodzi.

[6]. "Automatic Attendance System Using Face Recognition Technique", Mayur Surve, Priya Joshi, Sujata Jamadar, Minakshi Vharkate.

[7]. "Class Room Attendance System Using Facial Recognition System", Abhishek Jha

[8]. "Online Examination System", Firas A. Abdullatif

Social Distancing Violation Detector

Sumit Pradhan[1], Naman Gupta[1], Vibhooti Gonnade[1], Ms.Subhashree Rath[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore,
Karnataka, India

ABSTRACT

The ongoing COVID-19 corona virus outbreak has caused a worldwide disaster with its rapid deadly spreading. Due to the unavailability of effective antidote and the deficiency of immunizations against the virus, exposure of this virus to population increases. In the current situation, as there are vaccines available but not abundant; therefore, social distancing is thought to be an adequate precaution (norm) against the spread of this notorious virus. The risks of virus transmission can be reduced by maintaining an adequate distance among ourselves. The purpose of this work is to use Neural Network for social distance tracking using an overhead perspective. The framework uses the YOLO object recognition technique to identify humans in video sequences. Yolo uses a pre-trained algorithm that is connected to an additional trained layer using coco dataset. The detection model identifies peoples in sequence of steps. The distance between human is calculated using Euclidean Distance.

Keywords— Social Distancing, COVID-19, Machine Learning, Decision-Making, Deep Learning, Computer Vision, Training Data

I. INTRODUCTION

In February 2020, first three cases of covid-19 were detected in India. After this the government started imposing various restrictions on people travelling back to the country. It included 14 days of quarantine for people coming back. Government also advised people to maintain social distancing and other norms like sanitizing your hands and wearing face masks, because of these measures the country was able to contain the spread of the virus effectively.

But in 2021, country was hit by the second wave of Novel Corona virus and the number of cases increased drastically. Moreover, the virus mutated with some of its strains and the virus became deadlier than before with increase in mortality rate of people suffering from this virus.

After the first wave of corona virus people started taking the social distancing norm lightly resulting in the sudden surge of corona virus cases in 2021 also known as the second wave.

Fig 1. Social Distancing

II. SCOPE OF THE PROJECT

The system will be able to detect if people are violating social distancing. Presently the user will be providing pre-recorded video to system and based on that system will work.

To perform object detection, we use state of the art algorithm YOLO. YOLO is built in such way that it approaches over the image/video frame at once and a single convolutional neural network is deployed over the image/video and resulting in formation of bounding boxes for the classes with the label and its confidence and here we have our class as person. Computer vision using CNN (convolutional neural network) enables a computer to identify or detect process /object in image or video as humans do. Due to advances in AI and various innovations in deep learning and neural networks, this field has taken leaps in past years and has surpassed humans in some tasks of object detection and labelling.
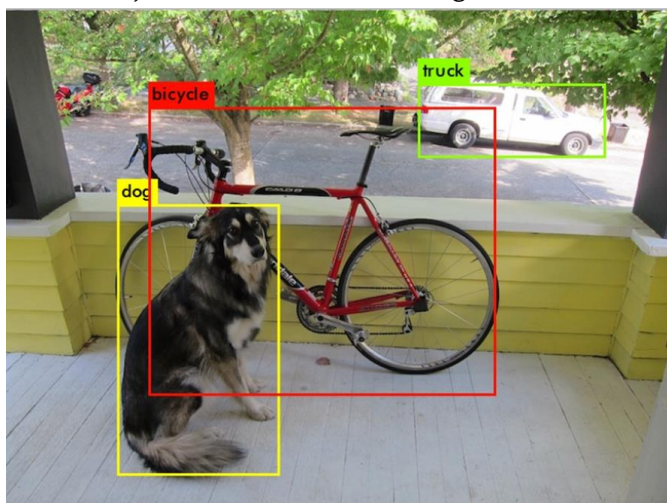
Fig 2. Object Detection using YOLO

III. PROPOSED SYSTEM

The Aim of this project is to detect the object in the images. and by understanding the relation between the object we can analyse the scene in the images. To perform this, we use YOLO v3. An object detection technique.

The object detection system is the most important aspect of this report. This is because of the part of this analysis that focuses on evaluating a person's position from the input picture.It is therefore necessary to choose the most appropriate object detection model to avoid any problems in the detection of individuals.

YOLO is state of the art object detection model which is fast and accurate especially when it comes to real time processing.

The system will take an image or video frame as input. The image will be forward propagation pass through the deep neural network as yolo takes the entire image in a single instance to make predictions.

Each image has a shape (608,608,3) as the framework is trained to run on 608x608 images. Further the framework divides input images into SxS grid, for each grid cell, consider a label y holding eight values as dimensional vector when we consideration person as our class.

· pc denoting the probability of class/object present in the grid or not

· bx, by, bh, bw specify the bounding box, bx, by are the centroid

· bh, bw as the ratio of respective height and width of bounding box to the height and width of the corresponding grid cell.

· c means the class, e.g. (c=1 is Person)

The class score calculation is done by $score\_c_i = p\_c * c_i$ Which means that the probability of presence of an object is p_c times the probability that the object is a certain class c_i.

After the calculation and prediction of score, box and class, the next step is the bounding box formation around the classes detected.

In each cell it can give 5 bonding boxes and the model is capable of predicting total of 19x19x5 = 1805 boxes with a single instance at the image. But for better results we required algorithm output with high accuracy, to achieve the we will be using NMS (Non-Max Suppression) to remove bounding boxes with less score or low confidence due to uncertainly of detecting a class or object.

Also overcoming the multiple bounding boxes overlapping for detecting the same class and get a single box in such case.

Fig 3. Yolo v3 Architecture
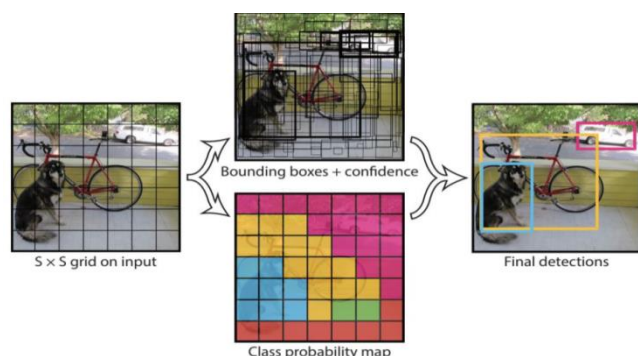
Fig 4. Processing of image using yolo

IV. REQUIREMENTS

HARDWARE REQUIREMENTS

- Processor: i5 Processor
- RAM: 8GB
- Hard Disk: 10 GB
- Any desktop / Laptop system with above level configuration or higher level.
- GPU: NVIDIA 1080 or NVIDIA Titan

SOFTWARE REQUIREMENTS

- Operating system: Windows 10
- Language: Python
  IDE: Python IDE, Anaconda, Google Colab

V. MACHINE LEARNING

Machine learning is a type of artificial intelligence that enables applications of software to become more precise in predicting results without being specifically programmed to do so. Various methods are used to measure the efficiency of ML models and algorithms. The trained model can be used after the conclusion of the learning process to identify, predict, or cluster new research data using the knowledge gained during the training process.

VI. DEEP LEARNING FOR COMPUTER VISION

Computer vision is an area that involves making "see 'a machine. Instead of the human eye, this device uses a camera and computer to locate, monitor and quantify targets for further processing of images. Example of image processing includes:

- Normalizing photometric properties of the image, such as brightness or colour.
- Cropping the bounds of the image, such as centring an object in a photograph.
- Getting rid of digital noise from an image, such as digital artifacts from low light levels.

With the evolution of computer vision, such technology has been widely used in the field of agricultural automation and plays a key role in its development. Deep learning in computer vision has made quick progress over a short period. Few of the implementation where deep learning is used in computer vision include face recognition systems, self-driving cars, etc. Convolutional neural networks aka convnets, a type of deep learning model used in computer vision applications. So, the aim of CNN is to perform two tasks: first is feature extraction and second is aggregating all the extracted features and making a prediction based on it.

VII. YOLO v3 (YOU ONLY LOOK ONCE)

YOLO, another way to deal with object recognition. Earlier work on article recognition repurposes classifiers to perform discovery. All things considered,

we outline object location as a relapse issue to spatially isolated jumping boxes and related class probabilities. A solitary neural organization predicts jumping boxes and class probabilities straightforwardly from full pictures in a single assessment. Since the entire location pipeline is a solitary organization, it tends to be streamlined start to finish straightforwardly on discovery execution.

Our bound together engineering is very quick. Our base YOLO model cycles pictures continuously at 45 edges for every second. A more modest variant of the organization, Fast YOLO, measures a shocking 155 casings for every second while as yet accomplishing twofold the guide of other constant finders. Contrasted with best-in-class location frameworks, YOLO makes more restriction mistakes yet is far more averse to foresee bogus recognitions where nothing exists. At last, YOLO learns general portrayals of items. It beats all other location techniques, including DPM and R-CNN, by a wide edge while summing up from common pictures to work of art on both the Picasso Dataset and the People-Art Dataset."

The entirety of the past item discovery calculations uses districts to limit the article inside the picture. The organization doesn't take a gander at the total picture. All things considered, portions of the picture which have high probabilities of containing the article. YOLO or You Only Look Once is an article recognition calculation vastly different from the district-based calculations seen previously. In YOLO a solitary convolutional network predicts the jumping boxes and the class probabilities for these crates

VIII.   SYSTEM ARCHITECTURE

The work has been developed using YOLOv3. First, we apply object detection algorithm YOLOv3 to get Humans. The bounding boxes obtained contain all the objects detected in the image belonging to dataset and it filters only the classes of persons.

- Data Source – open-source coco.
- Image Processing – ML Model consisting YOLO v3 using Darknet, Convolutional Neutral Networks.
- Results – Detection of Violation of Social distancing.
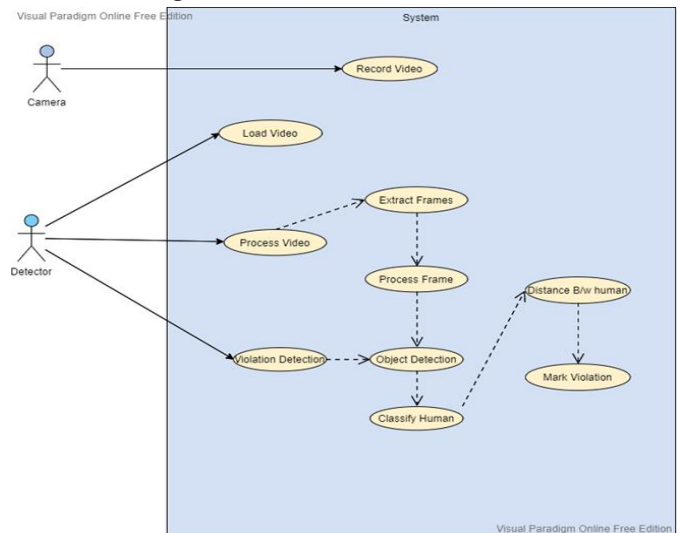
Fig 5.Use case Diagram

From the usecase diagram we can see there are two actors which interact with the system i.e. camera and the detector.

Camera records the video. The detector loads the video it then extracts frames, process it and find the objects in it. Later the system classifies humans from the objects and finds the distance between each human and if the distance between them is less than violation is marked.
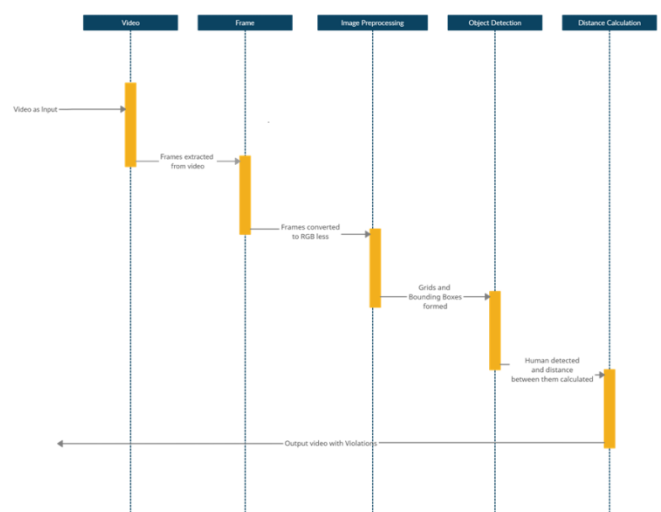
Fig 6. Sequence Diagram

Sequence diagram depicts the sequence of events that take place in the system.

IX. CONCLUSION AND FUTURE WORK

One of the essential precautions is minimizing physical contact that can lead to the spread of coronavirus is social distancing. Higher rates of virus transmission would be triggered by the effects of non-compliance with these guidelines. To incorporate the proposed features, a framework must be built using the Python and OpenCV Libraries. The key feature is the identification of social distancing violation's using yoIov3.

This system can be integrated with face detection in order to find the people who are violating the norm and fine them.

X. REFERENCES

[1]. Joseph Redmon,Santosh Divvala,Ross Girshick,Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection" [J].2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:779788.

[2]. Joseph Redmon, Ali Farhadi,"YOLO9000: Better, Faster, Stronger", The IEEE Conference on Computer Vision and Pattern Recognition

[3]. Rumin Zhang, Yifeng Yang, "An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera", 2018 Twelfth International Conference on Sensing Technology (ICST)

[4]. https://www.sciencedirect.com/science/article/pii/S2405959519304187

[5]. https://link.springer.com/chapter/10.1007%2F978-3-319-590721_10

[6]. https://ieeexplore.ieee.org/document/8310092

[7]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm

[8]. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[9]. https://iopscience.iop.org/article/10.1088/1742-6596/1821/1/012049

Customer Segmentation Analysis

Mr. Manjunatha Swamy[1], Payel Pattanayak[2], S. Jagan[2], Mohith Teppola[2]

[1]Senior Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Customer Segmentation Analysis is one of the business growth support systems. This is used to grow the business in the competitive market. This paper proposes an approach to develop a web portal based on target customers using predictive approach. After identifying the customers of same taste, they will be served products accordingly.

Keywords : Clustering, Customer Segmentation, Data Mining, Personalization, Ecommerce.

I. INTRODUCTION

With the evolution of new technologies and increasing growth of e-commerce it is important for every business to adapt new strategies which help them to win the competitive environment. Customer is the most valuable asset for any business. In this emerging market it is very difficult to maintain its customer base. To overcome this difficulty in a business has to focus on customer segmentation.

Customer segmentation means categorizing all the customers into same group. It is a technique in which we divide the customers based on their purchase history, gender, age, interest, etc. It is required to get this information so that the store can get help in personalize marketing and provide customers with proper deals. With the help of this project, companies can run targeted campaigns and provide user-specific offers rather than providing same offer to everyone. It acts as a base for Customer Relationship Management (CRM) for businesses to know the customer according to their transaction and focus on them separately. There are already various existing predictive models which provide info on customer segmentation and which helps them to segregate customers. For a successful business making more profit from each customer is key task apart from retaining and adding new customers. Many Different variety of models exist which helps the business managers to implement the strategic ways according to individual customer taste but each one with its own limitations.

II. RELATED WORK

Identifying right customer and providing right service at right time and treating different kind of customers differently is the key to run the business successfully. So, a predictive model will be used to distinguish customers into different groups based on the data collected. Once the customers are segregated then their behavioral buying pattern are noted to increase the profit for the organization future new customer.

Integrated Approach of finding customer segments along with their associative buying pattern is as shown below
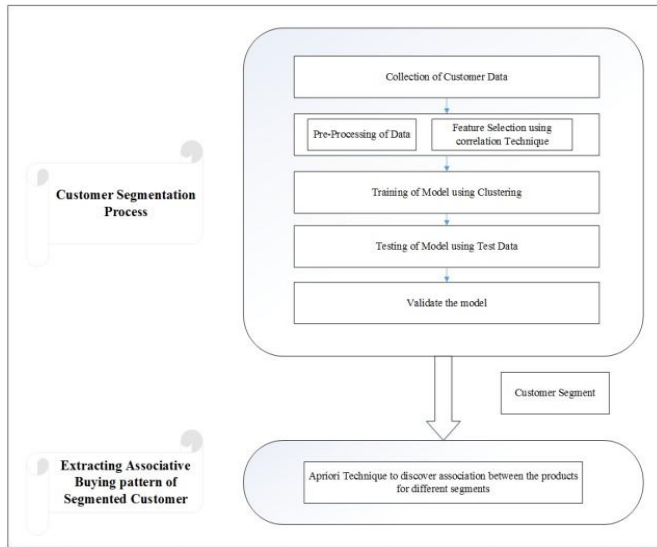
Fig. 1: Integrated Approach of finding customer segments along with their hidden buying patterns

Phase 1: Customer Segmentation

- **Step 1:** Collecting the Customer Data: It involves in the collecting the customer's transactional data that comprises of their statics (E.g.: Age, Gender etc.) and dynamic data (Eg: Purchase frequency etc.) from shopping vendors.
- **Step 2:** Pre-processing of Data: Pre processing of the data is one of the important steps for the accuracy of predictive model. In this step, the collected data will be segregated and required features will be extracted. Feature selection is responsible for extracting relevant features required for input vector of predictive model; it is a data reduction technique. This helps in as creating a subset of original features by not including the features which are redundant.

This paper proposes correlation technique helps in extracting the relevant features of the customers. Correlation measures the relationship between two features. It basically filters those features which are not required to form subset of the original features.

To measure the relation between the two features, the correlation coefficient is taken and is calculated between two features and based on its value. Correlation is broadly categorized into three categories as follows:

Positive correlation:

If two features are related in such a way that if one of them increases then the other also increases or if one of them decreases then other one will also decreases then this is called as positive correlation.

Negative correlation:

This correlation occurs when one decreases other increases.

No correlation:

This occurs if there is no relationship between the two features (i.e. the features are independent. So take in consideration of those features which are independent to form input vector for the model.

- **Step 3:** In this step we must pass the input vector to the model for training. After training, the model will then divide the data of the customer into homogeneous segments.
- **Step 4:** After the model is trained we should pass it to the test data to check its accuracy and efficiency.
- **Step 5:** After all this the predictive Model will now predict segments of future customer data.

Phase 2: Extracting the associative buying pattern of segmented Customers

Once the customers are segmented according to phase 1 we find associative buying pattern using association mining and appropriate technique from the particular segments to profit the business.

III. METHODOLOGY

The main purpose of the project is to analyse the customer based on various factors:-

- Age
- Gender
- Purchase history
- Wish lists
- Frequent visits
- Most searched
- Purchase range
- Purchase frequency

Based on the data collected, we categories the customers and understand their requirements. This helps us in satisfying customer expectations. There are mainly two benefits of the project, it helps us understand the customer and suggest them products that might be useful to the customer and eventually increasing the business. These suggestions play a big part in marketing the product.

The other benefit is, it helps us generate offers/coupon codes to the products that the customer desire to buy. The better we understand the customer the better we are able to generate offers. The ultimate goal is marketing, the better we understand our customers the better services we will be able to provide them. Further taking consideration to their search history we can analyze the customer more deeply and provide them with better search results.

There are four common types of segmentation:

1) Demographics
2) Geographic
3) Behavioral
4) Psychographics

Demographic: This is the most common type of segmenting the customer by considering parameter such as age, gender, occupation, income, marital status to create customer segments. It's much easier to get data for demographic segmentation than the other types of customer segmentation .We simply ask the customer to fill forms to get these data.

Geographic: This involves segmenting the customer based on their country, climate, state, region .In this segmentation we must keep the local culture and weather a factor for providing coupons and deals to the customer via messages. By doing so, it provides a greater value to consumer and encourages buying them.

Behavioral: This segmentation is implemented by the basis of how the customer interacts with your brands. This is a the category of segmentation which studies the behavioral traits of consumer, their likes /dislikes, they're response to the product they bought, the kind of brand they choose and promote.
This segmentation is mainly to understand the desire of the customer and to service those products to them.

Psychographic: Just understand ding the customer needs and interests wouldn't suffice, we must know the stage of the buying process they are in, i.e. their current social status. This segment is based on dividing the customers based on psychological factors, which includes behaviors, personalities, lifestyles and beliefs. This segmentation is mainly used to predict how the customer will respond to focused marketing campaigns.
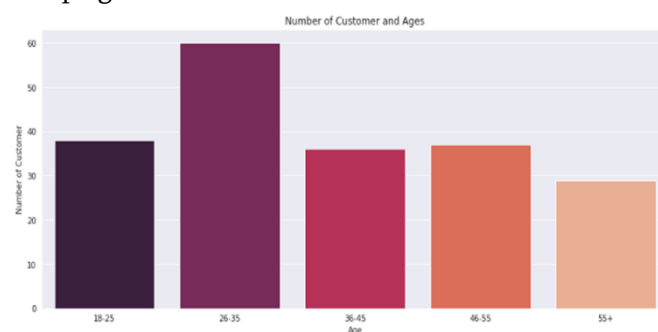
Fig. 2: Result Analysis

The K mean algorithm is discussed as follows:

K-Means is probably the most well-known clustering algorithm in machine learning. It is being taught in a lot of introductory data science and machine learning classes.

1. To begin, we first select a number of groups to use and randomly initialize their respective center

points. To identify the number of groups to use, it's good to take a quick analysis at the data and try to identify any distinct groupings.

2. The data point is classified by computing the distance between that point and each group center, and then classifying the point whose center is closest to the group center to be in the group.

3. Based on these classified points, we recomputed the group center by taking the mean of all the vectors in the classified group.

4. Keep repeating these steps for a set number of iterations or until the group centres don't change much between iterations. There is also an option to opt to initialize the group centres in a random way in a few times, and then select the run that has given best results which is expected.

On the other hand, K-Means has some disadvantages. Firstly, you have to select how many groups there are. This isn't ideal with a clustering algorithm we'd want it to figure those out for our analysis because the point of it is to gain some insight from the data collected. It may also starts with random choice of cluster centers and because of that it may give different clustering results on different runs of the algorithm. Due to this the results aren't repeatable.

IV. CONCLUSION

Customer segmentation is a way to understand and analyze the customer in order to provide customer satisfactory services, but the ultimate goal is satisfying customer so the company's business is maximum. We provide a variety of services like, product suggestion, unique offers to each individual, coupon codes, etc. We categorize each customer with a unique clustering algorithm. The data Is collected based on each individual search history, their wish list, etc. The ultimate goal is to help a company make a successful business

V. REFERENCES

[1]. Al-Qaed F, Sutcliffe A. Adaptive Decision Support System (ADSS) for B2C E-Commerce. 2006 ICEC Eighth IntConf Electron CommerProc NEW E-COMMERCE InnovConquCurr BARRIERS, Obs LIMITATIONS TO Conduct Success Bus INTERNET. 2006:492-503.

[2]. Mobasher B, Cooley R, Srivastava J. Automatic Personalization Based on Web Usage Mining. Commun ACM. 2000;43(8).

[3]. Cherna Y, Tzenga G. Measuring Consumer Loyalty of B2C e-Retailing Service by Fuzzy Integral: a FANP-Based Synthetic Model. In: International Conference on Fuzzy Theory and Its Applications iFUZZY.; 2012:48-56.

[4]. Magento. An Introduction to Customer Segmentation. 2014. info2.magento.com/.../

[5]. An_Introduction_to_Customer_Segmentation...

[6]. Baer D. CSI : Customer Segmentation Intelligence for Increasing Profits. SAS Glob Forum. 2012:1-13.

[7]. Kishana R. Kashwan, Member, IACSIT, C.M.Velu, "Customer Segmentation using clustering and Data Mining Techniques", International Journal of Computer Theory and Engineering, Vol.5, No.6, December 2013.

[8]. R.Kaur ,K.Kaur, "Data Mining on Customer Segmentation:A Review", International Journal of Advanced Research in Computer Science,Volume No-5,2017

[9]. Mark K.Y.Mak,George T.S.Ho,S.L.Ting,"A Financial Data Mining Model for extracting Customer Behaviour", INTECH open access publisher, 23 July 2011

[10]. LuoYe, CaiQiuru, XiHaixu,LiuYijun and Zhu Ghuangping , "Customer Segmentation for Telecom with the k-means Clustering Method", Information Technology Journal 12(3):409-413,2013 .

[11]. I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.

FoodDroid AR : A Wholesome Approach to Your Food Ordering Experience

Varun Rajesh Makhija[1], Vishnu Vardhan Vemuru[1], Shreyas Chandrashekar[1], Dr. Pamela Vinitha Eric[2]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Augmented Reality creates an experience using the user's real-time environment that allows them to interact with it. Today's rapidly developing technologies have made smartphones, laptops, and wearable devices more accessible, thus enabling AR-based applications.

Keywords — Augmented Reality, 3D Models, FoodDroid AR, AR Core, Photogrammetry

I. INTRODUCTION

This paper will focus on two things, the first being the problem that FoodDroid AR attempts to solve, and the second being how it solves the problem.

The food ordering and delivery industry has consistently been growing over the years; however, the customer experience hasn't evolved much since the advent of such services. While ordering, people tend to do either of two things: one, they order food items they are already familiar with, or two, they order new (From the customer's perspective) food items on an experimental basis by judging the items in question via a picture (If it exists) or a description (If it exists). We are going to focus on the second aspect wherein many times, you would find neither a picture nor a description for some food items, and upon ordering them, you may not be satisfied with either the food item itself or the quantity of the food item that you received. FoodDroid AR aims to solve this problem using the marvel that is Augmented Reality. By making use of 3D rendered models of the food items available within food ordering and delivery app services, FoodDroid AR allows the user to view any food item they desire in the form of a fully rendered 3D model and can also choose to project said model onto their real-time environment by making use of their Android phone's camera system. This would help the user interact with their order more efficiently but also help them gauge the quantity of food that would be delivered to them.

II. REQUIREMENTS

A. Hardware Requirements:

Android mobile device: An ARCore supported Android Nougat 7.0 with API level 24 or higher is required to run Scene Viewer to display AR models that are generated. The AR models are displayed within the environment using the mobile camera, which is accessed by Scene Viewer.

GPU: A CUDA-enabled Nvidia GPU is required for Meshroom to generate AR models.

B. Software Requirements:

1) ARCore: Google's platform for building augmented reality experiences. ARCore enables your phone to sense its environment, understand the world and interact with information. ARCore uses three capabilities to allow virtual content to interact with the real world as specified in [1]. They are:

- Motion tracking allows the phone to understand and track its position relative to the world.
- Environmental Understanding allows the phone to detect the size and location of all types of surfaces: horizontal, vertical, and angled surfaces like the ground, a coffee table, or walls.
- Light Estimation allows the phone to estimate the environment's current lighting conditions.

2) Java: Java is a general-purpose class-based, object-oriented programming language that is intended to let application developers write once and run anywhere while having as few implementation dependencies as possible. The language is computer architecture-free, and the applications created are typically compiled to bytecode that can run on any Java Virtual Machine as mentioned in [2].

3) Meshroom: Meshroom is a software developed by the AliceVision Association that uses a collection of images to create 3D models using a technique called Photogrammetry. In order to perform said task, Meshroom requires an NVIDIA Cuda-enabled GPU with CUDA-10 compatibility and compute capability between 3.0 to 7.5.

4) Scene Viewer: An immersive viewer used by an Android application to facilitate 3D and AR experiences. Many Google partners have implemented it to support AR reliably. FoodDroid AR uses Scene Viewer. to allow the food AR models created on Meshroom. Details about Scene Viewer are in [3].

5) Android Studio: It is the official integrated development environment for Google's Android Operating system. It was jointly developed by JetBrains and Google. The IDE supports Kotlin, but here the preferred language of choice is Java.

III. MAKING THE 3D MODELS

Making 3D models can be an extremely complex process; however, using Photogrammetry technology, we were able to create 3D models of food from scratch. Photogrammetry is used to create a map or 3D model of real-world objects or scenes from a collection of photographs of said object/scene. We used AliceVision's Meshroom, a software specifically designed to render 3D Models from photographs using Photogrammetry. An example of using Meshroom for photogrammetry can be found in [4]. The process of creating a 3D model of a food item, as followed for FoodDroid AR, can be broken down into x steps. They are as follows:

A. The Photography Environment:

1) The room in which you click the photographs must be as clutter-free as possible. This means that you keep only the essential items in the room: no unnecessary real-world objects or scenes must be there as this will confuse the software while building the model.

2) An ideal case scenario would be a room with a single table for the food item to be placed and absolutely nothing else in the room. If the room has white walls and flooring, that will only enhance the quality of the model being rendered.

B. The Environment Lighting:

1) Once the room has been set up, it is imperative that the lighting is even across the object from all angles.

2) The ideal case would be to have a powerful light source on top of the table that would light up the object appropriately.

C. Clicking Pictures of the subject:

1) Meshroom supports almost any smartphone camera, as most smartphones encode picture metadata within the photographs themselves.

2) Hence, Meshroom can identify things like the camera aperture for depth of field, the ISO for picture granularity, and the shutter speed for sharpness.

3) It is essential to avoid any motion blur and depth blur, which is why it is ideal to use a tripod stand for the smartphone.

4) Finally, use the smartphone camera to click at least 60 pictures of the subject by going 360 degrees around the subject and getting pictures from eye level as well as above-eye level of the subject.

5) The more the pictures, the more details will be rendered in the 3D model.

D. Importing the pictures into Meshroom:

1) The photographs need to be imported into Meshroom, which is a pretty straightforward process: you can drag and drop.

2) Once imported, save the project as a Meshroom file and head over to the workflow.

E. Rendering the 3D Model:

1) You can alter any settings in the workflow if you would like, concerning MeshFiltering, Meshing, etc.

2) In order to begin rendering the model, head over to the Texturing tab and right-click and click compute.

3) The colours on each tab in the workflow indicate the state of the computation. Blue signifies that it has been submitted for computation, Red indicates an error, Orange indicates the computation is going on, Green indicates that the computation is completed.

F. Exporting the results:

1) The final model will be a Textured Mesh that will be saved as an OBJ file with the corresponding MTL and Texture files.

G. Converting to GLTF 2.0:

1) Since SceneViewer can only display 3D models in GLB/GLTF 2.0 format, we must convert the OBJ file to the required format.

2) Any online 3D Converter can convert the OBJ file and corresponding MTL and Texture files to a single GLTF 2.0 file.

3) This file can then be hosted to display in the FoodDroid AR app.

IV. DISPLAYING MODELS IN AN AR ENVIRONMENT

Scene Viewer from Google's ARCore is used to display the models in an AR environment. The main reason for choosing Java as the preferred language of development instead of Kotlin is because Scene Viewer can only be used in Java. Scene Viewer requires the Google Play Services for AR in order to display the models in AR. Scene Viewer has certain strict model requirements, and they are as follows:

- The file format must be GLB or GLTF 2.0.
- The model size should not be over 10 MB.
- It is recommended not to bake shadows for models as Scene Viewer creates hard shadows for the models.
- The maximum texture resolution must be 2048x2048.

In FoodDroid AR, all the AR models used are hosted on GitHub so they can be loaded by Scene Viewer using a link provided to the model's raw content.

Scene Viewer is launched using an explicit intent on Android. It requires three different intent parameters as follows:

- A file URI must be passed in order to load the AR model into Scene Viewer.
- A title for the model should also be passed.
- A mode parameter must also be specified. The mode parameter allows Scene Viewer to determine which mode the model must be loaded in. The modes are as follows:

A. 3D preferred:

Displays models in a 3D viewer by default but has a button specified to allow AR mode.

B. 3D only:

Displays models in a 3D viewer only and does not allow an AR viewer.

C. AR preferred:

Displays models in an AR viewer but has a button specified to allow 3D mode.

D. AR only:

Displays models in an AR viewer only and does not allow a 3D viewer. This mode is generally discouraged as a mobile may not support AR mode.

The mode used in FoodDroid AR is 3D preferred since it is safer to load a 3D viewer in case a mobile phone does not support AR. The model can then be viewed in AR mode by clicking the view in your space button. Scene Viewer allows resizing, moving, and zooming into the models using finger gestures to increase the user experience. Optional parameters like sound, link, and resizable can also be used but are not used in this project.

V. CONCLUSION

The food ordering and delivery industry is an evergreen one that is constantly growing. In order to make a mark in the vast scope of this industry, one is required to give more value to the user with innovative solutions. While AR is yet to enter the food industry completely, a few restaurants have employed an AR menu experience for diners. The process to generate models for each of the food items in a food delivery app is a tedious process, but the right amount of workforce can turn it into a much smoother process. The results we saw in our application proved to represent the original food item quite well, although it does not appear exactly as in real life. Eventually, the aim should be to make the food items look as authentic as possible and make them available as a possible feature in food delivery applications.

A. Benefits:

1) FoodDroid AR is a food ordering app that allows you to view food items in a 3D format.
2) FoodDroid AR can also project these 3D models in real-time environments.
3) FoodDroid AR aims at reducing complaints of any sort and thus improving customer satisfaction.
4) 3D models projected by the android system AR core are used to show the customers exactly what they pay for or how much quantity they get.

B. Disadvantages:

1) Due to the limitations of pre-existing android smartphones, not all of them can render and view 3D models.
2) Rendering and projecting 3D models need a stable internet connection and suitable system hardware.
3) To be able to build 3D models, you require powerful software, which may not be cost-effective.

VI. REFERENCES

[1]. "ARCore overview | Google Developers", Google Developers, 2021. [Online]. Available: real-world objects or sceneshttps://developers.google.com/ar/discover. [Accessed: 08- May- 2021].

[2]. H.Schildt, "Java: The Complete Reference, Eleventh Edition," McGraw Hill.

[3]. "Using Scene Viewer to display interactive 3D models in AR from an Android app or browser", Google Developers, 2021. [Online]. Available: https://developers.google.com/ar/develop/java/scene-viewer. [Accessed: 08- May- 2021].

[4]. "Tutorial: Meshroom for Beginners — Meshroom v2021.0.1 documentation", Meshroom-manual.readthedocs.io,2021. [Online]. Available: https://meshroom-manual.readthedocs.io/en/latest/tutorials/sketchfab/sketchfab.html. [Accessed: 10- May- 2021].

Smart Remote Classroom

Nagarjun U[1], Katherine Sanjana. L[1], Deepthi Reddy GB[1], Sivabalan N[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Road, Panattur post, Kadubeesanahalli, Bengaluru-560103, Karnataka, India

ABSTRACT

With the increase of technology and with the increase in the strength of the students and the number of the departments in the educational institutions, it is difficult to exchange the study materials between the students and faculties. And also because of the pandemic in the recent time many students are facing problems to attend classes. Humans being visual creatures. What we learn visually will be retained better than something that is read out to us. This why the technology for the smart classes as become a boon for the students in the 21st century. Therefore this project will help students to learn the concepts visually. On the other side it helps the faculty to take the attendance with the help of face recognition system. Also the classes will be recorded and stored into the drive for the future use or to help the students who don't have a proper internet facility. There may be a problem for the students to take down the notes hence this project will provide speech to text conversion technique for the purpose of notes. This will work as an application where we can login in to the live class with ID for the class and password. The main goal is to allow the students to attend for the classes virtually and learn the concepts very clearly and provide video and audio recordings and also notes for those students who are facing internet issues.

Keywords— Smart Remote classroom, face recognition system, e-learning, online learning

I. INTRODUCTION

Online Learning has spread a wide range of technologies such as the email, chat, new groups and texts, worldwide web, audio and video conferences delivered in computer networks for imparting education. It helps the learners for learning at their own pace, according to one's own convenience. Online Education requires a great range of resources and a careful planning. In this, teachers will be acting as facilitators rather than transmitters to the content of knowledge, and ICT is regarded as the resources that enhanced the learning experience for the students. Learners learn by e-learning tools which is available to all. E-Learning has been brought back for the joy in learning by its innovative and interactive of content delivery and has proved that to be more appealing among the students E-learning is something which is growing very rapidly in these time since it helps us to attend to the classes even in these pandemic time. In this project we will be implementing some extra concepts like face recognition and Speech to text conversion and we can record audio and video lectures of the class. All these are the extra concepts included in the project. The chat box option will also be provided in the project. With the increase of technology and with the increase in the strength of the students and the

number of the departments in the educational institutions, it is very difficult to exchange the study materials between the students and faculties. Given our current pandemic situation we have to come up with a solution that helps the students to gain education in an effective manner so that we prevent an education gap in our student's lives.

II. LITERATURE SURVEY

Literature survey has become the most important step in the process of software development.

The e learning applications are the technology that are offering the learner-cantered opportunities will redefined concept of learning organization. Why the expertise in the soft technology will be given trainers new credibility; why time will not spend will be the scarce resource for learning competes with the other organizational demands; the valid of corporate universities and also the virtual business schools. And what one be learnt from different strategic responses to the e-learning of the blue-chip companies [1].

The complex interaction between the e-learning and the institution and the people attitude, a more strategically approach is necessary to ensure that the e-learning has best possible chance get to succeed [2]. Open and the Distance Learning Series has mention .This book has provided a series of proven and practical guideline for using the technology in the education, giving readers the overview of how technological applications to education can be developed [3]

Face recognition and utilizing the technology to map a unique pattern to develop a desired outcome would be an emerging topic as an active research area driving many disciplines, such as the image being processed with respect to the pattern recognition and neural networks [4]. Face recognition technology has got a huge response and many commercial and the law enforcing applications developed in the recent years. An automatic system based on pattern transforms for face recognition [4]. Face recognition

technology is getting a huge response and numerous commercial and law enforcement applications developed in recent years. An automatic system based on pattern transforms for face recognition.

Speech to text conversion:

The aim of the particular review understands how STR technology is being used to support learning over past 15 years, and the second is for analysing all research evidences for understanding Speech-to-Text Recognition technology that can enhance learning [5]. The findings have discussed from the different perspectives : (a) potentials of the STR technology, (b) its use by specific the groups of users from different domains, (c) quantitative or qualitative research methodologies being used, and (d) The STR technology implications [5].

Chat box

In the recent decades, internet forms an indispensable trend through making resources of learning and research ready obtainable for both students and teachers allowing them to exchange and share information [6].

The learning based on computer involves a complete framework of hardware and software usage that generally allow the obtainability of information and communication technology. However, the engagement of computer is applied for purpose of saving and recapturing information in order to manage education [6].

III. COMPARATIVE ANALYSIS

In this paper the comparative analysis mainly done based on the time, space and the relevancy. Found in the previously existed system.

The E-learning platforms which are present now are have many features like session record which helps in recording the present session for the future need There will be a chat box where we can send or receive messages that will help us communicate with

each other. Another feature of the E-learning platform is screen sharing which is very needful for teaching or explaining the concept for the students and in the same way even students can share the information with everyone else also these are the some of the features of the existing E-learning platforms.

Disadvantages:

➢ Few requirements like gadgets, internet connection are essential, but not everyone is affordable.

➢ Technology generates noise interference causing disturbance.

➢ Many ideas which are illogical to implement using technology at present as well as in future

IV. PROPOSED SYSTEM

The proposed system is having some other extra features like Audio and video record which helps in recording the session for future use.

- We can convert the Speech to Text which helps to get the notes even if are not able to note down the points.

- There will be a chat box where we can send or receive messages that will help us communicate with each other.

- Design the system structure that should provide 24/7 availability and we have been keeping those features in our mind that the structure should be flexible enough to support all the time and will be available throughout the day which gradually extending the idea of the Smart Classroom. Step

Advantages:

- The face recognition system can help the teacher take attendance easily. It will reduce their effort to take attendance by just doing face recognition.

- Speech to text recognition helps the students to get proper notes even if they are not able to take notes.

- Audio/video recorder helps to record the session which can be used further.

- This will help in providing effective remote education to all.

This will also help the students who don't have a proper internet connection all the time in their by providing video and audio recordings which will be saved in the drive and can be accessed by the student any time after the class.

V. ARCHITECTURE OF THE SYSTEM

The similar separation of the pattern recognition algorithms to four groups is proposed by the colleges. We can grope face recognition methods to three main groups.

The following are the proposed approaches:

- Templates matching: - Patterns are represented by the sample, curves, and textures. The recognition function is usually the correlation or the distance measure.

- Statistical approach: - Patterns are being represented as the features. This recognition function is the discriminant functions.

- Neural networks. This representation may differ. There is network function in the some point

Workflow

The conditions for how the secure virtual should be conducted has been mentioned in the points below:

➢ User id should be long enough and should be unpredictable, for both the students and the teachers

➢ Check if User Id is being valid and allocates the session.

➢ Check if User id being generated by any unauthorized automation.

➢ Session id can be scheduled by the teacher and can be expired after some period of time or else

also when the teachers privilege has been changed use the cookies to display session details and IDs. Avoid persistent login options (such as the remember me option while login); session expire on the security error like not authorized User or if there is any illegal actions.

➢ Face recognition for student Authorization. Remove session cookie when a session is cancelled or destroyed.

➢ In face-to-face sessions, every faculty will have his or her own method of teaching. Each varies from one to one in style and is susceptible to the mistakes. You can eliminate the issues with e-learning. Online learning provide   every time. Each learner will go through the same experience regardless of when or where he or she takes the course.

➢ The above points describes that how secure is the e-learning platform is and how will the platform work is also described in a step by step manner in order to conduct a secure and safe online or also called as the virtual class

Fig 1 system architecture

Here in the diagram it displays all the layers of the system how the workflow is done.

VI. IMPLEMENTATION

The flowchart diagram below clearly describes the working of face detection algorithm which is used for the project
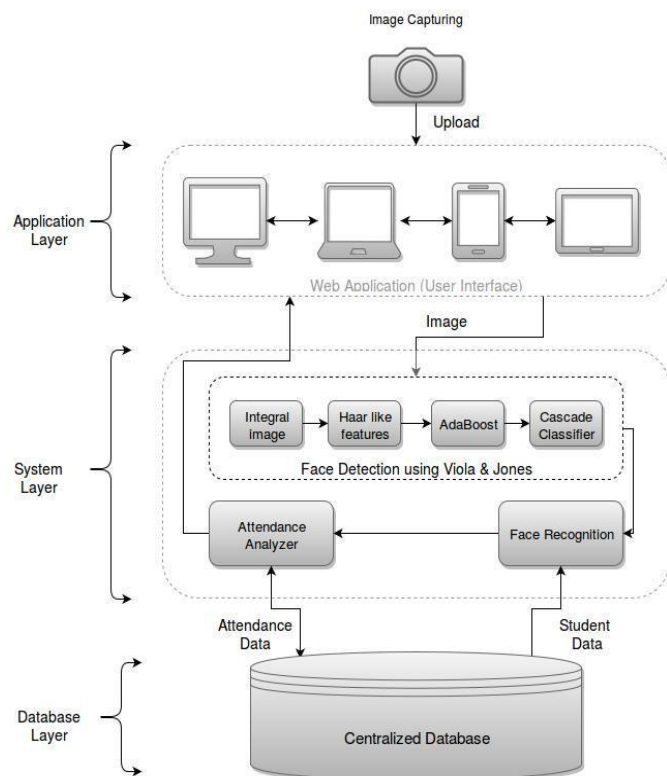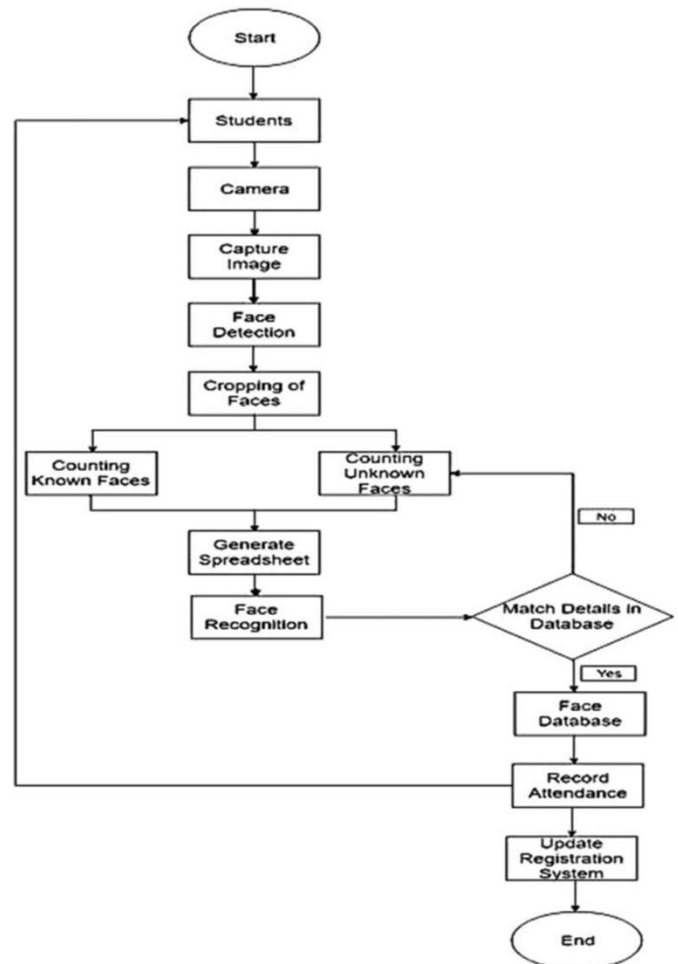
Fig 2 face recognize system flow chart

VII.   FUTURE WORK

Humans being visual creatures. What we learn visually will be retained better than something that is read out to us. This why the technology for the smart classes as become a boon for the students in the 21st century. Therefore this project will help students to learn the concepts visually.

The more interactive way of learning or teaching: Instead of watching to diagrams on a blackboard as chalk drawn, students can now experience subject in a high-definition on the smart screen, making the learner and learning interactive and the fun. Teachers can use the interactive module and visually appeal method to teach like videos, virtual reality and presentations to help their students better in grasping the subject. Immersing the students with the subjects that help them to learn and retaining subject better. After all, textbook can be fun to read, but nothing can be compare seeing the jungles of the Amazon is the high definition video to seeing a picture of it in the book.

Helping for those who are struggling with the traditional learning methods: Not all the students will enjoy the chalk and board teaching methods and it can be very difficult to capture the attention and make students learn. The videos and the other interactive modules of smart remote classroom will work better at capturing the imagination and students attention. Smart classes are also very beneficial for the students with learning disability as they easily grasp what is taught by the medium of video and interactive modules.

VIII. CONCLUSION

Using a E-learning is something which is growing very rapidly in these time since it helps us to attend to the classes even in these pandemic time. In this project we will be implementing some extra concepts like face recognition input word and Speech to text conversion and we can record audio and video lectures of the class. All these are the extra concepts included in the project. The chat box option will also be provided in the project.

With the increase of technology and with the increase in the strength of the students and the number of the departments in the educational institutions, it is difficult to exchange the study materials between the students and faculties.

Given our current pandemic situation we have to come up with a solution that helps the students to gain education in an effective manner so that we prevent an education gap in our student's lives.

Given our current pandemic situation we have to come up with a solution that helps the students to gain education in an effective manner so that we prevent an education gap in our student's lives.

Aim of project was to design the system structure that should provide 24/7 availability and we have been keeping those features in our mind that the structure should be flexible enough to support all the time and will be available throughout the day which gradually extending the idea of the Smart Classes.

IX. REFERENCES

[1]. M. Sloman, "The E-learning Revolution," The E-learning Revolution, CIPD,, 2001.

[2]. P. A.Amira, "Automatic face recognition system," P.ferrel ,ECIT, Queen"s University.

[3]. N. N. u. Rustam shadiev, "The Speech-to-Text Recognition Technology for Enhancing Learning," Speech-to-Text Recognition Technology for Enhancing Learning.

[4]. A. A. I. &. G. R. Jaffal, "chatbots' impact on the studen communication," chatbots' impact on the studen communication , 2009.

[5]. A. Z. Ayham Fayyoumi, "The Novel Solution Based on the Face Recognition to Address Identity Theft and Cheating in Online Examination Systems," Advances in Internet of Things, vol. 4, April 28, 2014.

[6]. B. &. M. J. Collis, "The Flexible learning in a digital world.," Flexible learning in a digital world.London, 2001.

Corona Killer A Free Augmented Reality Game Using Unity

Savion Mario Sequeira, Dinesh G

Department of Computer Science and Engineering, New Horizon College of Engineering, Outer Ring Rd, near Marathalli, Kaverappa Layout, Kadubeesanahalli, Bengaluru, Karnataka, India

ABSTRACT

Corona Killer is an Augmented Reality game developed for Android devices using Unity 3D Engine and Google AR Core. This paper will focus on the process and the technologies used in order to achieve the game.

Keywords — Augmented Reality, Unity, Android, Game Development, XR Development

I. INTRODUCTION

There are two major components introduced in more detail in this paper, Augmented Reality Development and Game Development.

"Augmented reality combines the physical world with a computer-generated or simulated environment. It is accomplished by superimposing computer-generated images over real-world images. There are four forms of augmented reality: marker-based, marker-less, projection-based, and superimposition-based. It has a wide range of real-world applications. AR is used in a variety of areas, including medicine, education, manufacturing, and robotics, to name a few. "[1]

The way I would define Augmented Reality is bringing the digital word into the physical space with the means of a device, the device being a smartphone in this scenario. It is augmenting the digital world to the real world.

Augmented Reality is a subset of XR (Extended Reality), this includes Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR). Mixed Reality combines both Virtual Reality and Augmented Reality.

"Augmented Reality has come a long way and has recently been popularized in mainstream entertainment through games like Pokémon GO and Face filters.

Game Development is a very complicated process, in reference to the book "Agile Game Development with Scrum" by Clinton Keith, he mentions about how the pioneer days of game development has now disappeared. It is no longer a sole programmer who used to specialise in every aspect of game development from programming and even rendering their own art. They have all been replaced by an army of specialists. "[2]

Game Development is a very vast process, it goes beyond the scope of just being a programmer or an artist, I classify game development as a multidisciplinary field. From your basic programming and scripting, to animation, 3d modelling, game design, level design, sound design and so much more. It has become a very complicated process. It can take an experienced game developer a year or more to develop a full-fledged game from scratch for production.

This should not scare away any solo developers trying to get started or create a game. The beauty of game

development is how huge the community is, there are millions of free and paid assets which anyone can use under certain conditions for their game. As a result if I as a developer is good at game scripting I need not worry about learning to make good 3D models anymore as I can just download them. This is why game development is so accessible in general to anyone who is willing to give it a go.

Unity 3D game Engine, is one if not the most commonly used game engine by solo and even some company game developers. Unity has a huge resource online from their own sources and even from third party sources. It is a very well documented game engine which makes it easier for anyone new trying to get started.

The reason I used Unity is because it supports both Game Development as well as Augmented Reality Development. It is suitable for any cross-platform development without having to change the codebase for each platform. Speaking of codebase Unity can be scripted in C# or JavaScript. There are other ways to do scripting for people who are weak in coding, using UI drag and drop to create scripts.

II. REQUIREMENTS

The requirements for developing a game in AR is surprisingly almost nothing. Let's start with AR. When it comes to developing in XR, VR can get very expensive to develop fir, starting with getting the equipment required to handle VR which can cost a lot of money, and moreover the equipment required to handle VR is also expensive. As a result for anyone dipping their toes in XR development and does not currently have the resources to build for VR, they can always develop for AR.

The main requirement for AR development is a smartphone. Smartphones nowadays are built to handle augmentation. Any Android device running Android 7 or above can support AR. This is because firstly Google Play AR services is only compatible with Android version 7 or above. Moreover, the new

smartphones come with depth sensors and this is very important since any 3D object to be Augmented requires all three axis coordinates to be spawned into. This being the x, y and z. A normal camera takes a picture or records a video in 2D, but depth sensors allow the camera to now differentiate between an object in the foreground and an object in the background. From just the AR perspective these are the only requirements to allow a device to use augmentation.

When it comes to the developing phase things can get a bit expensive depending on the scale of your project. Firstly, you need a laptop with around 16 GB of RAM to use Unity without any hassle, Unity is a heavy application since it is very extensive. The next question would be do you require a GPU? Well from a game that does not depend on too many heavy graphics you will not require a GPU as your PC does not need to render any heavy graphics, you can use a few 3D models and effects and get away with a good CPU. Saying this the moment you start adding a lot of graphics into your game be in in AR or normal games you will start to face lagging issues; hence GPUs are a must in those scenarios.

To end the requirements chapter, I would say compared to some of the other project you can start on, AR development is very easy to get it done and get started with for even a student with almost no budget.

III. SETTING UP YOUR WORK ENVIRONMENT

"The Unity Hub is a stand-alone programme that makes finding, downloading, and managing Unity projects and installations much easier.

Unity will prompt you to instal the Hub if you start the editor without it installed. You can generate an empty project from the Hub installation prompt if you have a licence from a previous version of Unity. "[3]

Unity is very easy to setup, as mentioned earlier it is very well documented engine. As a result almost any

issue you might face will have been resolved. To get started all you need to do is download a version of Unity from the Unity website. Unlike other platforms you can download multiple versions of Unity each being stored as a separate application. This is because some versions have LTS (Long Term Support) while the newer ones are still in Beta. Moreover, some versions are compatible with some features like AR while some might not be, this makes developing very flexible as you can choose which version of Unity suits your needs and which version of Unity you are comfortable developing in.

When downloading you can add several modules that you will require for your project , in this case you can add the Android Module which includes JDK and NDK tools that is required for creating an APK to run on your smartphone.

Once everything is installed you can now create a new project, a 3D project with the version of Unity you require. The last step you need to do to allow development in AR, is install Unity AR foundation and Google AR core. Let's look at each individually.

"Within Unity, the AR Foundation allows you to work with augmented reality platforms across several platforms. This kit provides an interface for Unity developers to use, but it does not provide any AR functionality. To use AR Foundation on a target system, you'll also need separate packages for the Unity-supported target platforms:
AR Core XR Plugin on Android and AR Kit XR plugin on iOS" [4]
AR Core is the official plugin created for allowing AR applications to run on Android devices. It is not exclusive to Unity; it can also be used in Android Studio.

To install these packages, you can download both AR Foundation and AR Core from Unity's package manager. Once installed you can begin developing an AR application.

IV. BRIEF SUMMARY OF THE GAME

Corona Killer is a 3D AR shooter, essentially the game requires you to move your phone around finding viruses and destroy them by vaccinating them. When the player downloads the game , they are first shows a home screen where you can see how to play the game, about menu, and even an achievements menu. In a later chapter we will talk about the importance of minor details that can make a game exciting or boring to play.

The main button is the Play Button which the player will click to start the game, once the player clicks that the game begins. After a small delay the viruses start spawning randomly in various places, using a cross hair on the screen the player will need to point at that virus and click the vaccinate button. The more points the player scores the harder the game gets, there is a health bar to add stakes to the game, once the health reduces to 0 the player loses. To add advantage to the player, random heart points are also spawned in between the player on shooting them gains back some health. Once the player dies, the high score is checked. The goal of the game is to beat the previous high score. In addition to that, the player can also gain some achievements to make it more interesting. This is a brief summary of the game, we will now look at the game development process.

V. APPLYING SCRUM TO THE GAME DEVELOPMENT PROCESS

"A game is a type of software designed to entertain people. However, in real-world game development, merely following the software development life cycle (SDLC) is insufficient, as developers face numerous obstacles during the life cycle. To resolve the problem, game development employs a method known as the game development life cycle (GDLC) to guide game development. None of the current GDLCs, on the other hand, specifically discuss how to produce a high-quality game. This paper introduces a new game development life cycle model as well as recommendations for delivering a high-quality game.

At each point, a number of quality parameters are specifically considered." [5]

With reference to the above paragraph, I realised there was no particular way to go about a game development process, so you are required to create a workflow which works for you, especially when you are doing it solo. You can then look back at the process and refine it for your next project until you have a very functional workflow.

For this project I followed a SCRUM approach and broke down my project into separate components tacking each component one at a time. Below was my entire thought process for this project.

- Figure out the main game mechanic of being able spawn objects in AR at random intervals. To do this we can us an IEnumerator.
- Next would be to figure out if our phone is pointing at the spawned AR object in free space, this can be solved using RAYCASTING. We will look at Raycasting with more detail in Paragraph VI, as it is the crux of this project.
- Once we have figured this out, we need now a point which can be used to point at the objects, we can create a cross hair on the center of the screen now. This cross can be referenced to act as the point for applying Raycasting to detect the object.
- Now that we are able to detect the objects, we can now start performing an action when the object is detected, to do this we need a trigger, that trigger can be a button at the side of the screen.
- Once the button is clicked and the objected was detected we can now destroy the object and perform a random effect.
- Now that the main part of the game is done, we need to start adding game mechanics, first is a score system.
- The score system will give players a reason to play the game, a sense of achievement. Each time an object is destroyed we can increase the score by one.

- Now that the player can achieve something we need to start adding a risk factor, in this case a health bar, this health bar gives the player a reason to play well. If the virus on screen starts spreading, we then need to start reducing the health overall. Until the player eventually dies.
- Now that we have implemented the health system, we need to allow the players to gain some health back, we can start generating hearts for players to collect.
- With this we are done with the game mechanics. We can now start working on UI components. We first need to make the main screen which the user will access.
- The main screen will have multiple components as well, starting with the Play button.
- Then we can add an Options button to control the game volume.
- We can now add a help menu to make the player understand the concept of the game.
- Lastly it is important to give goals for the player to achieve so we will add an achievements menu.
- We can also add an about menu to give some information about the developer of the game.
- To end the game, we need a game over menu, with option to retry and quit the game. We add this after the player loses all health.
- The last feature to show if the player improved from last time, we can add a high score, which only gets updated once the player beats his previous score.
- With this we have developed the entire game. We can refine it by adding some sound effects and music to enhance the overall experience.

VI. RAYCASTING

Raycasting is the method of firing an invisible ray from a point in a specific direction to see if any colliders are in its way.

This is a simple definition of raycasting but this technique is what allowed me to develop my entire game.

"AR technology, which shows virtual objects overlapping in the real world, is gaining popularity, and a variety of AR applications are being created. These apps were first created as mobile AR for mobile devices including smartphones and tablet computers. Mobile AR systems used the smartphone's camera to display the real world while augmenting simulated objects on the camera screen. Since then, smartphones have become more widely available, making AR applications accessible to all. With the development of computer vision technology, the floor and wall can be distinguished by a mobile device with Apple ARKit or Google AR Toolkit [6] and the furniture can be placed in the desired position without markers. The most significant disadvantage of AR based on mobile devices is that the user must continue to keep the device in order to superimpose the virtual entity, which is the most significant limitation of AR based on mobile devices." [7]

The reason I cited the above article is because to show the current flaw behind the system, raycasting will only work if the player moves the phone to the location to detect an existing object in front of it.

The description of raycasting in unity is "Casts a ray, from point origin, in direction direction, of length maxDistance, against all colliders in the Scene." [8]

Basically, when you point your phone at an object, the application shoots a beam to the object and if the beam collides with the object a hit is detected. We can configure for only certain object to be detected and then perform any action of those objects when detected. It shoots a beam from the centre of the phone.

In my game I have named the virus as "Cube" and any clone created of the virus would be called "Cube(Clone)" so if the player points the crosshair at the virus and click the vaccinate button, the program first checks if the object detected is named "Cube(Clone)" and if so then I programmed the object

to be destroyed while displaying a special effect and incrementing the score counter.

This is the main technique behind which the game works, luckily in Unity we do not have to worry about coding the entire technique as Raycasting is available as a function to which we can input parameters to it.

While saying this , it has been pointed out in several papers about the flaws of Raycasting and new techniques are being developed to improve this based on Raycasting itself. For my project it worked well and I did not face any issue using this technique.

VII. CONCLUSION

The future of gaming is very fruitful and bright. The market is ever growing and there seems to be no stop to its growth. As estimated to hit staggering numbers of over 250 billion dollars in the coming years, game development is extremely sustainable for any engineer willing to become a part of the gaming industry.

Technology is ever growing, withing a year the technology can become obsolete since new and better ways are being developed. The amount of game developers are also rising with time, due to the ease of accessibility of the game engines being developed such as Unity and for sure in the coming years they will become more powerful and more easier and efficient to use by anyone willing to learn and become a game developer.

The current project being worked upon can be improved upon vastly in the coming years with better 3D rendering upgrades being developed and more powerful processors and graphical cards being made, it will be very easy to quickly render projects and test them. It will enable developers to try to make more complex games and execute more interesting ideas due to this, as they would not have to spend years working on a single project.

Augmented Reality has leaps and bounds to go further as it is still in its early stages. Augmented

Reality which is a subset of Extended Reality is collectively being developed by vast number of companies. In this field particularly Facebook is investing heavily to make more complex interpretations of what Augmented Reality and Virtual Reality currently can do. The technology will become even more accessible to the developers and more consumer friendly in the coming years. The technology needs a lot of improvement in how functional the portability of the sensors can be.

Currently the smartphones are being developed with sensors and hardware components equipped to handle augmented reality applications. While they work pretty well, the detection at times can still be very clumsy with irregular surfaces and it can cause glitches in the application.

This project has ways to go and can be improved in several ways, with more time available and with better hardware equipment the time taken to make a more complex game can be greatly reduced.

VIII.   ACKNOWLEDGMENT

IX. REFERENCES

[1]. Riya Aggarwal; Abhishek Singhal, "Augmented Reality and its effect on our life", 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019

[2]. Clinton Keith, "Agile Game Development with Scrum", Chapter 1

[3]. The Official Unity Webpage for getting started with Unity. https://docs.unity3d.com/Manual/GettingStarted.html

[4]. The Official Unity Webpage for getting started with XR Foundation https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.1/manual/index.html

[5]. R. Ramadan and Y. Widyani, "Game development life cycle guidelines," 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2013, pp. 95-100

[6]. Bellarbi, A.; Zenati, N.; Otmane, S.; Belghit, H.; Benbelkacem, S.; Messaci, A.; Hamidia," A 3D interaction technique for selection and manipulation distant objects in augmented reality.", n Proceedings of the 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B), Boumerdes, Algeria, 29–31 October 2017; pp. 1–5.

[7]. Ro, H.; Byun, J.-H.; Park, Y.J.; Lee, N.K.; Han, T.-D., "AR Pointer: Advanced Ray-Casting Interface Using Laser Pointer Metaphor for Object Manipulation in 3D Augmented Reality Environment.", Appl. Sci. 2019, 9, 3078.

[8]. The Official Unity Webpage for getting started with Raycasting https://docs.unity3d.com/ScriptReference/Physics.Raycast.html

Prediction of Crop Yield and Cost by Finding Best Accuracy using Machine Learning Approach

Swathi[1], Mrs. Soja Rani[2]

[1]M Tech Student, Department of CSE, New Horizon College of Engineering, Bangalore, Karnataka, India
[2]Senior Assistant Professor, Department of CSE, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Among around the world, agribusiness has the significant duty regarding improving the financial commitment of the country. Still the most agrarian fields are immature because of the absence of arrangement of biological system control advances. Because of these issues, the yield creation isn't improved which influences the farming economy. Subsequently an improvement of rural profitability is upgraded dependent in plant yield expectation. Forestall issue, Agricultural surroundings need for anticipating the yield for given particular data set utilizing AI procedures. The outcomes show that the visibility of the proposed AI calculation strategy can be contrasted within best exactness with accuracy.

Keywords— Dataset, Machine learning, Random Forest Decision Tree, Prediction of Crop

I. INTRODUCTION

In agricultural nations, cultivating is considered as the significant well spring of income for some individuals. In current years, the rural development is locked in by a few advancements, conditions, methods and civic establishments. Furthermore, then usage of data innovation may change the state of dynamic and in this manner ranchers may yield the most ideal way. For dynamic cycle, information mining procedures identified with the horticulture are utilized. Information mining is a cycle of separating the hugest and helpful data from the gigantic measure of data sets. . Plant development expectation is proposed for checking the plant yield adequately through the AI strategies. It is additionally relevant for the

computerized cycle of cultivating is to start another time that will be reasonable tor the ranchers tor specialists to take proposal about the fitting yield on explicit area of their property and don't have any desire to fail to remember any progression of the development all through the cycle harvest created in harvest or plant yield forecast since agribusiness has distinctive information like soil information, crop information, and climate information Despite the fact that, the sentiment from specialists is the most advantageous way, this application is intended to give precise arrangement in quickest way conceivable. This present exploration's fundamental goal is to bring cultivating measure a bit nearer to the advanced stage.

II. LITERATURE SURVEY

A writing audit is an assortment of text that survey the basic purposes of current information on and additionally methodological ways you deal with a specific subject. It is optional sources and talk about distributed data in a specific branch intelligent and once in a while data in a specific branch of knowledge inside a specific time span. Its objective is to carry the per user in the know regarding flow writing on a theme and structures the reason for another objective, for example, examination that might be required for the territory and goes before an exploration proposition and might be only a straightforward rundown of sources. Generally, it has a hierarchical example and joins dual synopsis and combination. A synopsis is of significant data about the source, yet a combination is a re-association, reshuffling of data. It may give another translation of old material or consolidate new with old understandings or it may follow the scholarly movement of the field, including significant discussions. Circumstance, the writing survey may assess the sources and exhort the per user on the most appropriate or pertinent of them.

The work examined[1] Fast assurance of soil natural issue (SOM) utilizing relapse models dependent on soil reflectance ghastly information serves a significant capacity in accuracy agribusiness. "Deviation of curve" (DOA)- based relapse and incomplete least squares relapse (PLSR) are two displaying ways to deal with anticipate SOM. In any case, not many examinations have investigated the precision of the DOA-based relapse and PLSR models. The DOA-based model, which requires just 3 groups in the noticeable spectra, likewise gave SOM assessment satisfactory precision. The outcomes demonstrated that both displaying strategies gave sensible assessment of SOM, with PLSR beating DOA-based relapse all in all. Be that as it may, the presentation of PLSR for the approval data set diminished all the more perceptibly.

The examination [2] precisely surveying the hefty metal pollution in harvests is essential to food security. This examination gives a technique to recognize hefty metal feelings of anxiety in rice utilizing the varieties of two physiological capacities as separation records, which are acquired by digestion of distantly detected information with a yield development model. Two pressure records, which relate to every day all out issue change consolidated into the World Food Study (WOFOST) crops development model and determined by acclimatizing the model with leaf region file (LAI), which was brought for time-arrangement information. The feelings doesn't consistent with rice development hence, to improve the dependability, the two pressure tracks were gotten at both the first and the last half times of rice development.

In the paper [3] streamlining and implementation of a bordering field weight soil dampness sensor utilizing the printed circuit board intervention. It incorporates the investigation of a novel configuration of an inter digital sensor for estimating soil dampness with two existing configurations. The improved plans were mimicked by utilizing a 3-D finite-component technique and created by utilizing a copper clad board. The exhibition to manufactured sensors was assessed utilizing four soil tests gathered from various areas. The perceptions were contrasted with the standard strategy with assess the dirt water substance of the examples. The portrayal technique and the aftereffects of the entire detecting framework are talked about regarding alignment, dynamic test, and repeatability.

Paper analysis [4] which is a productive observational technique for arrangement and forecast is another way to deal with crop yield functions. It portrayed the corn yield assessment in Iowa State utilizing four AI approaches, for example, RF (Random Forest), ERT (Extremely Randomized Trees) and DL (Deep Learning). Likewise, examinations of the approval

measurements among those were founded. To look at the occasional sensitivities of the corn yields, three period bunches were set up: (1) MJJAS (May to September), (2) JA (July and August) and (3) OC (ideal mix of month). In general, the DL strategy indicated most elevated correctness's as far as the connection coefficient for the three time frame gatherings. The correctness's were generally great in the OC gathering, shows partial ideal blend of month can be critical in measurable displaying of harvest yields.

Examines the study of [5] Primary Study of Soil Available Nutrient Simulation Used for Modified WOFOST Model and Time-Series Remote Sensing Observations The methodology of utilizing multispectral distant detecting (RS) to gauge soil accessible supplements (SANs) has been as of late created and shows promising outcomes. This strategy conquers the constraints of usually utilized strategies by structure of a measurable design that associates RS-based harvest development and supplement content. In any case, the security and exactness of this design based require improvement. In this article, we supplanted the factual model by incorporating the World Food Studies (WOFOST) structure and time arrangement of detecting (T-RS) perceptions for guarantee strength and precision. Information was absorbed into the WOFOST structure to extrapolate crop development reenactments through a solitary highlight an enormous zone utilizing a specific digestion technique.

III. SYSTEM ARCHITECHURE

Investigate your information and you should not perform such a large number of activities in every cell. One choice that you can take with this task is to do a ton of investigations in an underlying journal. These don't need to be coordinated, yet ensure you utilize enough remarks to comprehend the motivation behind each code cell. At that point, after you're finished with your examination, make a copy note pad where you will manage the abundance and put together your means so you have a streaming, strong report and ensure that you keep your per user educated on the means that you are taking in your examination. Follow each code cell, or each set of related code cells, with a markdown cell to depict to the per user what was found in the previous cell. Attempt to create it with the goal that user would then be able to comprehend what they will be finding in the accompanying cell.
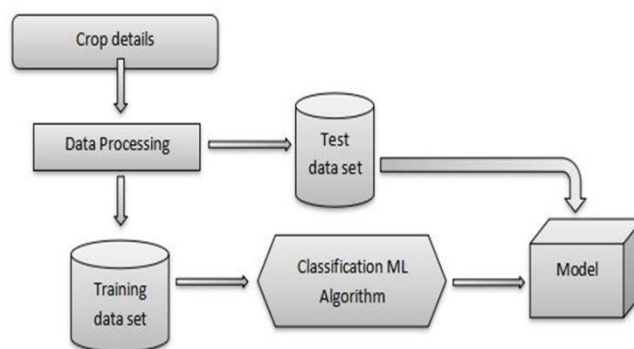
Figure: 1System Architecture of Crop Yield

Our objective is push for helping ranchers, government utilizing our forecasts. Every one of these distributions state they have shown improvement over their rivals yet there is no article or public notice of their work being utilized basically to help the ranchers. In the event that there are some real issues in revealing that work to next stage, at that point recognize those issues and have a go at settling them.

IV. DATASETS

A mix of models is a dataset and when working with AI procedures we normally need a couple datasets for various purposes.

A. Preparing Datasets

The demo dataset is right now gave to AI model dependent on this educational assortment the model is readied. Each new detail involved at the hour of design goes probably as a test instructive assortment.

After the movement of testing, model estimate reliant on the inferring it closes dependent on the readiness educational assortments. Satellite Imagery (Remote Sensing Data), has been comprehensively used for predicting crop yield. This dataset is accumulated using the sensors mounted on satellites or planes, which recognize the energy (electromagnetic waves), reflected or diffracted from surface of the earth. Inaccessible distinguishing data has a lot of energy gatherings to bring to the table, basically relatively few of them have been used for crop conjecture. In any case, there are a couple of gathering who have made a pass at creating material features using the gatherings which are typically dismissed, and they have been productive with improving results with that.

TABLE I

Variable	Description
Crop	Crop name
State Name	Indian state name
District Name	District name list of each state
Cost of Cultivation	Cultivation amount for C2 Scheme
Cost of Production	Production amount for Scheme
Yield (Quintal/Hectare)	Yield of crop
Crop year	Crop year list
District Name	District name for each state
Area	Total area of each place
Rainfall	Water availability of each crop
Average Moistness	Straightforwardly impacts the water relations of plant and by implication influences leaf development
Mean Temperature	Climate of r each particular crop

B. Exploratory Data Analysis

In this piece of the information, you will stack in the data, to Check for cleanliness and a short time later trim and clean your dataset for examination. Guarantee that you file your methods carefully and legitimize your cleaning decisions.

C. Training Dataset

The first line imports iris instructive assortment which is as of now predefined in sk-learn module. Iris enlightening assortment is basically a table which Contains information about various combinations of Iris blooms.

1. For model, to import any computation and train_test_split class from sk-learn and numpy module for use in this program.
2. This procedure secludes dataset into getting ready and test data discretionarily in extent of point we encapsulate any computation.
3. In the accompanying line, we fit our planning data into this estimation so PC can get readied using this data. By and by the readiness part is done.
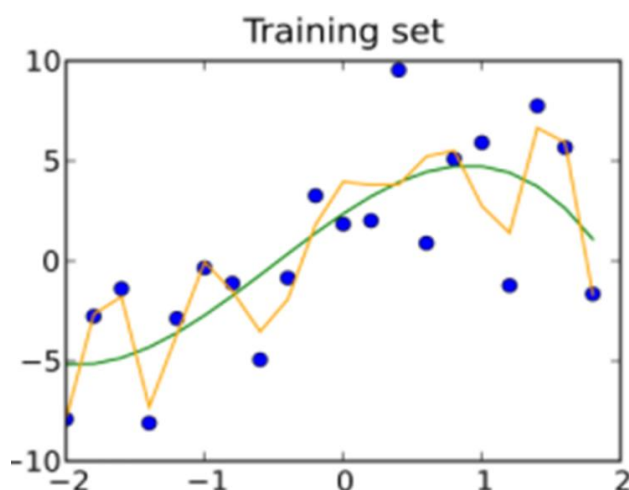
Figure 2: Training Dataset

D. Testing Dataset

1. Now we have measurements of another bloom in a numpy cluster called 'n' and we need to foresee the types of this blossom.

2. We do this utilizing the foresee technology which accepts exhibit info which lets active objective incentive as yield.

3. Objective worth recovers out to be 0. At last discovery grade which is the respective no. of expectations discovered right complete forecasts made.

4. This utilizing the technique that fundamentally analyses an real estimations of the test set with the anticipated qualities.
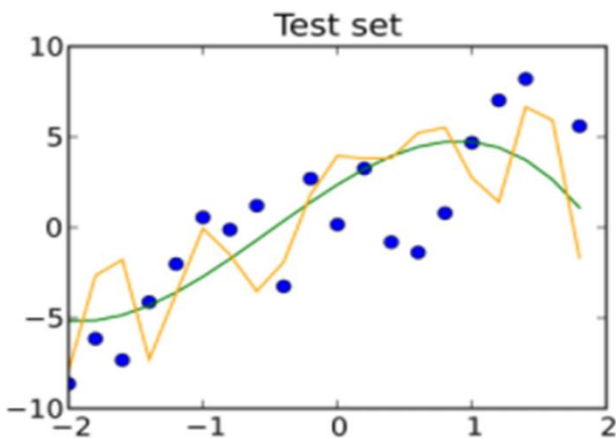
Figure 3: Testing Dataset

V. MACHINE LEARNING

Machine learning is concerned with learning the computer programs and improves automatically with the experience. However there is no clear idea about how to make computers learn as we humans do but there are several algorithms that are used for certain type of learning task. From age old year's humans worked under every different categories as time passed machines came into use which was trained with algorithms and spoon fed how to work then complete a given task and now it is the decades of the concept ML where in machines are not trained how to work in real but allowed itself to learn from experience using the input, analyze the data by itself and give the results.

Figure 4: Example of healthcare

MI has practically unlimited applications in the medical care industry. Nowadays AI is assisting with smoothing our own authoritative cycles in emergency clinics, guide and treat the Irresistible infections and customize clinical medicines.

A. ALGORITHMS

1. Random Forest

Random forest is an collection for decision tree mainly used for classification and regression. It is supervised algorithm that doesn't over fit the model, handles the missing value and mainly modelled for categorical value. It is better than decision tree because it contains only the subset of features. It is a classification algorithm consisting of many decisions trees.
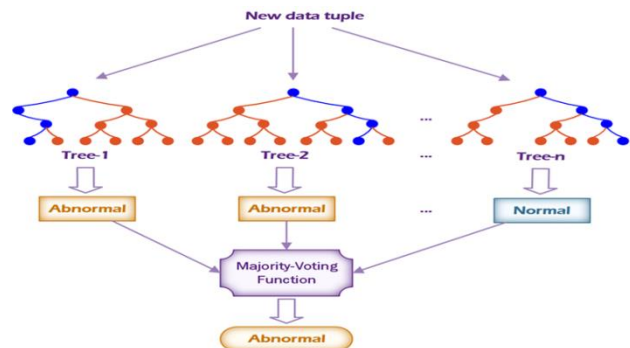
Figure 1: Random Forest

2. Decision Tree

The most frequently used supervised learning algorithm is decision tree. It works well for both categorical (discrete) as well as for continuous dependent variable. In simple words the structure is similar to a tree present in data structures. The tree consists of a main node Root internal nodes which is nothing but attributes given in input labelled based on some given conditions(test on the attribute) and finally the last nodes are the leaves which is the target concept / possible outputs for a given problem statement, branches are the values obtained after the attribute is tested.
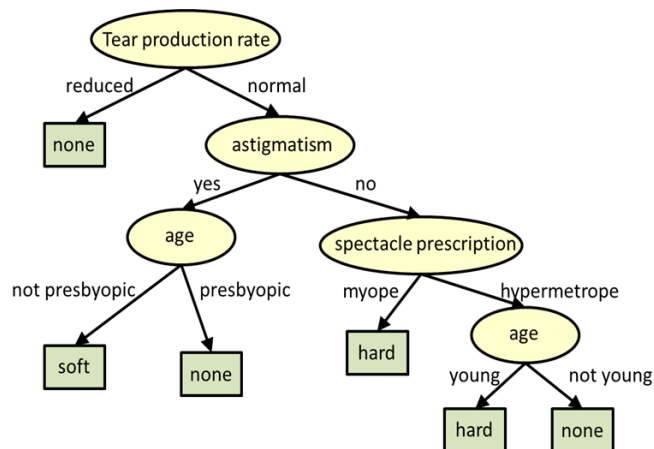
Figure 2: Decision Tree

3. Support Vector Machines

It is a support learning algorithm. In this modular the points   will be represented in N dimensional space so that examples with different features will be clearly separated by a large gap. The new instances will be Then classified to the category in which there is small gap. SVM can easily perform both linear and nonlinear classification using kernel trick that is it maps the inputs to high dimensional output. Nonlinear SVM is nothing but the boundary that the algorithm calculates is not a straight line. SVM is very much suitable in the following cases:

- When there is huge number of training data.
- When the number of zero values are more.
- To solve problems like image classification, Gene's classification etc.
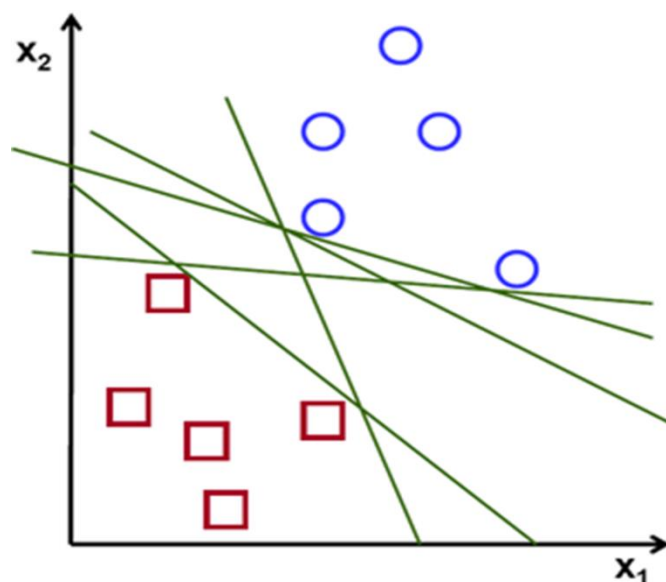
Figure 3: Support Vector machine

4. K-Nearest Neighbours

It is commonly used to solve the classification and regression problems. KNN algorithm will store all the available cases and when a new case is encountered the case will be assigned for class which is pure common among the K nearest neighbour calculated by Distance function. There are various distance functions such as Euclidian, Manhattan and Murkowski used for similarity of function and hamming distance used for particular sector variables.
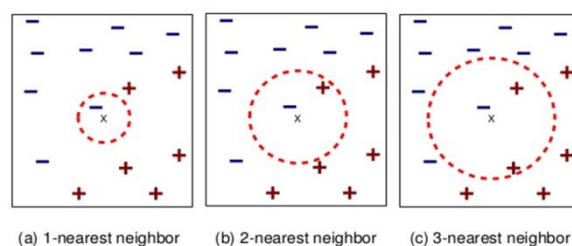
Figure 4: K-Nearest Neighbour

5. Logistic Regression

Logistic Regression is a AI calculation where it is utilized for the arrangement issues which is a prescient examination calculation and also dependent

through basic idea of portability. The hypothesis of regression tends it to limit the cost function between 0 and 1 , true or false, etc. Logistic Regression is a critical AI calculation since it can give probabilities and group new information utilizing nonstop and discrete datasets. Regression can be utilized to characterize the perceptions utilizing various sorts of information and can undoubtedly decide the best factors utilized for the arrangement.
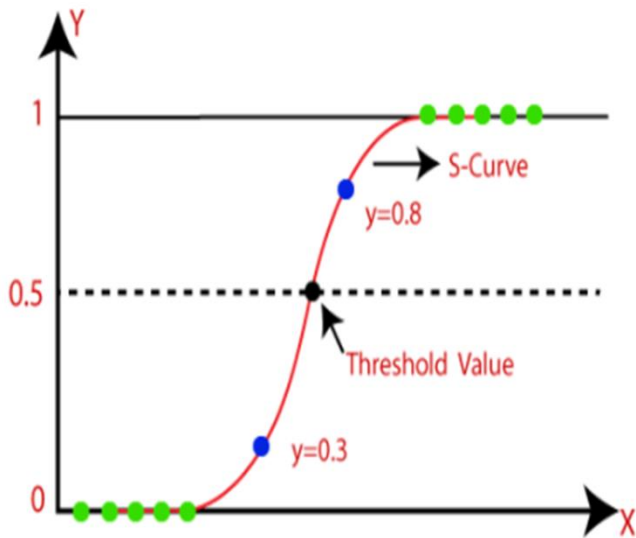
Figure 5: Logistic Regression

Below is an example logistic regression equation:

$$y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$$

Here y is the consisted yield, b0 is the tendency or else catch term and b1 is the coefficient for the single information regard (x). Each part in your data has a connected b coefficient (a predictable veritable worth) that ought to be acquired from your arrangement data.

VI. RESULTS AND ANALYSIS

Data collection is the route toward get-together and assessing information on components of interest. Information Cleaning implies the way toward recognizing the wrong, fragmented, erroneous, unessential or missing piece of the information and afterward changing, supplanting or erasing them as indicated by the need..

Now the data looks clean and organized, but dropping some of the columns such as State name, District Name, Season, Crop, that would be use to the analysis.

1. **Classification report of Logistic Regression Results**:
Accuracy of Logistic Regression is: 94.75627137970353
Confusion Matrix result of Logistic Regression is:
 [[41994 1646]
  [2033 24487]]
 Specificity: 0.9233408748114631

2. **Classification report of Decision Tree Classifier Results:**
Accuracy of Decision Tree Classifier is 100.0
Confusion Matrix result of Decision Tree Classifier is:
   [[43640   0]
     [02652   0]]
  Specificity: 1.0

3. **Classification report of Random Forest Results:**
Accuracy of Random Forest is: 100.0
Confusion Matrix result of Random Forest is:
 [[1776   0]
 [0 1087]]
Specificity: 1.0

4. **Classification report of Support Vector Machines Results:**
Accuracy of Support Vector Machines is: 85.78414250785889
 Confusion Matrix result of Support Vector Machines is:

[[1591 185]

[222 865]]

Specificity: 0.795768169273229
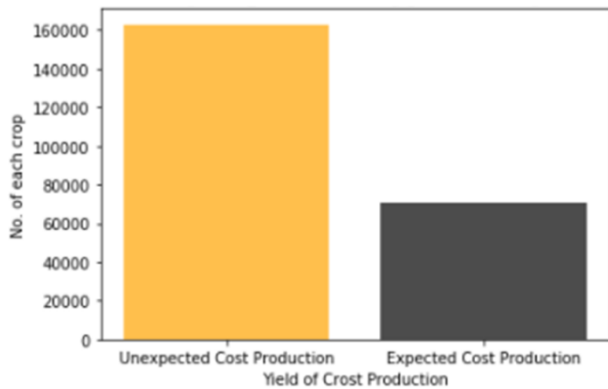
Figure 1: Prediction results expected by farmers by yield cost production amount.
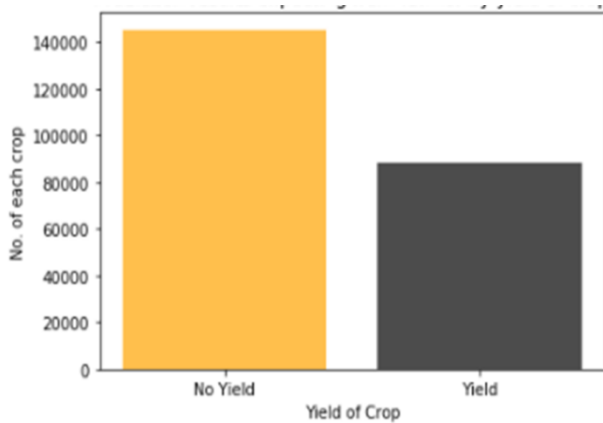
Figure 2: Prediction results expected by farmers by yield of crop.
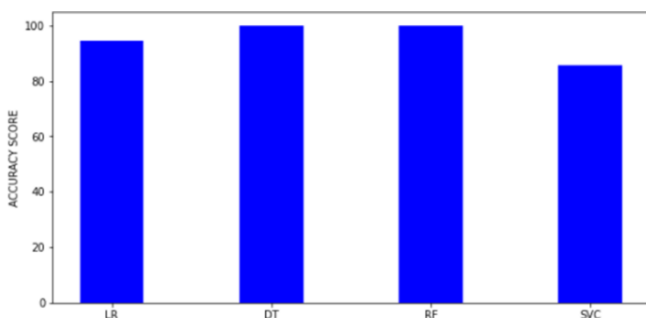
Figure 3: Best Accuracy Score Comparison

VII.   CONCLUSION AND FUTURE WORK

This paper works with analytics process started from data cleaning and taking care of, missing data, exploratory assessment finally model design and appraisal.. Finally we predict the crop using machine learning algorithm with different results. This brings a segment of the going with encounters about crop assumption. As most limit sorts of yields will be covered under this structure, farmer may get familiar with about the reap which may never have been created and runs through each and every conceivable yield, it helps the farmer in powerful of which gather to create. Similarly, this structure examines the past making of data which will help the farmer with getting understanding into the premium and the cost of various yields in market. Agricultural division needs to computerize the distinguishing the yield crops from qualification measure. To computerize this cycle by show the expectation brings about web application or work area application. To upgrade the work to execute in Artificial Intelligence climate.

VIII.   REFERENCES

[1]. G.L. Chmura, S.C. Anisfeld, D.RM. Cahoon and J.C. Lynch, "Global carbon sequestration in tidal saline wetland soils", Global Biogeochemical Cycles, vol. 17, no. 4, 2003.

[2]. W. Song, B. M. Chen and L. Liu, "Soil heavy metal pollution of land in China", Res. Soil Water Conserv., vol. 20, no. 2, pp. 293-298, 2013.

[3]. U. S. Lakshmi, D. N. Singh and M. S.V Baghini, "A critical review of soil moisture measurement", Measurement, vol. 54, pp. 92-105, Aug. 2014.

[4]. USDA(2012), Census of agriculture, United States Department of Agriculture, https://www.agcensus.usda.gov/ (last date accessed: 17 August 2016).

[5]. Huang, J.; Sedano, F.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Tian, L.; Zhang, X.; Fan, J.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agric. For. Meteorol. 2016, 216, 188– 202.

[6]. Kannan, M. Prabhakaran S and P. Ramachandran (2011).Rainfall forecasting using

data mining technique. International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401

[7]. Khan Mohammad A.,Md. Zahid-ul Islam and Mohsin Hafeez (2011). Evaluating the Performance of Several Data Mining Methods for Predicting Irrigation Waterrequirement. Proceedings of the Tenth Australasian Data

[8]. B. Abishek; R. Priyatharshini; M. Akash Eswar; P. Deepika Prediction of effective rainfall and crop water needs using data miningtechniques 2017

[9]. Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada 2019

[10]. Sheena Angra; Sachin Ahuja Machine learning and its application 2017

[11]. Filippo Sciarrone Machine Learning and Learning Analytics: Integrating Data with Learning 2018.

[12]. Pavan Patil1, Virendra Panpatil2, Prof. Shrikant Kokate Crop Prediction System using Machine Learning Algorithms 2020

[13]. Md. Tahmid Shakoor, Karishma Rahman, Chakrabarty.2017."Agricultural Production Output Prediction Using Supervised Machine LearningTechniques".978-1-5386-3831-6/17/$31.00 ©2017 IEEE

[14]. I. Ahmad, U. Saeed, M. Fahad, A. Ullah, M. Habib-ur-Rahman, A. Ahmad, J. Judge Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad Punjab

[15]. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system fo r precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 32-36. doi: 10.1109/ICoAC.2017.7951740

[16]. Naive Bayes classifier available at https://en.wikipedia.org/wiki/Naive_Bayes_clas sifier

[17]. S Brunda1, Nimish L2, Chiranthan S2, Arbaaz Khan2 Cro p Price prediction using Random Forest and Decision Tree Regression 2020

[18]. s://medium.com/swlh/random-forest-and its implementation-71824ced454f

[19]. Y. Peng, C. Hsu and P. Huang, "Developing crop price forecasting service using open data from Taiwan markets," 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)

[20]. Ananya Roy; Prodipto Das; Rajib Das Temperature and humidity monitoring system for storage rooms of industries 2017

[21]. Gerard Biau * GERARD.BIAU@UPMC.FRLSTA & LPMAUniversite Pierre et Marie Curie – Paris VI ´Boıte 158, Tour 15-25, 2eme etage ´4 place Jussieu, 75252 Paris Cedex 05, France

[22]. international journal of computer sciences and engineering 6(10):74-78 october 2018

[23]. Journal of Education and Practice www.iiste.org ISSN 2222-1735 (Paper) ISSN 2222-288X (Online) Vol.7, No.25, 2016

[24]. Support vector machine based machine learning method for GS 8QAM constellation classification in seamless integrated fiber and visible light communication system September 2020

[25]. A New K-Nearest Neighbors Classifier for Big Data Based on Efficient Data Pruning February 2020

Tribal Welfare Application

Harshitha H[1], Lokesh Divvela[1], Abhishek L[1], Pavan Kumar[1]

[1]Department of Computer Science, New Horizon College of Engineering, Kadubeesanahalli, Marathahalli, Bangalore, Visvesvaraya Technological University, Karnataka, India

ABSTRACT

Adivasi is the collective term for tribes of the Indian subcontinent. Tribes in India are not very up-to-date with the latest news and updates. So, there is a good chance of them missing out on the welfare schemes available for their benefit provided by state government and central government. Our aim is to make a simple mobile and web-based application allowing them to get news and updates focused solely on them. A mobile application to give information to tribes about welfare schemes of the ministry of tribal affairs and link it to various web portals of the ministry.

This app aims to be a one stop shop for all the tribal community helping them get information about all the latest schemes as and when announced by the Minister of Rural Development. Then prime objective of the app is to enhance the reach of welfare schemes of Government and fill in the gaps in service deficient tribal areas, in the sectors such as education, health, drinking water, agro-horticulture productive, social security, etc. Another priority for us is to keep the UI as simple as possible to make it easy for even new users of mobile phones and apps.

Keywords - Tribal Welfare, Welfare Schemes, Android Application, Schemes

I.   INTRODUCTION

In today's world communication has become so easy due to integration of communication technologies with internet. However the people living in villages and people who are not very fond of latest technology, find it very difficult to utilize it because of the fact that using them requires basic knowledge of mobile phones.

The community of people we are hoping to help here are the tribals. Adivasi is the collective term for the tribes of the Indian sub-continent, who are considered indigenous to places within India where they live, either as foragers or as tribalistic sedentary communities.

This paper aims at developing an app that will help tribals to learn about the services and welfare schemes benefitting them, all in one place. The system will be made with very simple UI. Ease of use will be of at-most importance.

II.   LITERATURE REVIEW

THANAL – App made for Kerala Government

At the annual startup conference held by the Kerala Startup Mission in 2017, the IT Secretary M Shivshankar had asked startups to come up with apps

that would provide the public with civic solutions. According to Ashik, Thanal collects information regarding user through just seven questions. "It categorises users for deserving welfare schemes, initiated by the Kerala government. Further, it provides the list of schemes the user is eligible for and gives details and processes of each one." he said.

Sarkari Yojana/Schemes – Launched by Govt.

Sarkari Yojana/Schemes launched by the government caters to all the categories of the citizens. However, Government schemes are mainly launched for the welfare, upliftment and assistance of the poor and weaker section of the society.

Gram Samvaad

Gram Samvaad is a citizen-centric mobile app, created by Ministry of Rural Development in collaboration with NIC team of Department of Rural Development, to serve and empower the rural citizens of India, by facilitating single window access by citizens to information at Gram Panchayat level on various Rural Development programs, covering inter-alia programme objectives, scope and performance.

III. PROPOSED SYSTEM

We are proposing to develop a mobile application with a very simple and minimal UI making it easy to use even for a newcomer. We want to develop this app keeping in mind Indian tribes. The app should be a one stop shop for all the Indian tribes to get information regarding all the available government schemes benefitting them, so that they don't miss any scheme and can make the most of what the government is offering them.

Our app can have a welcome screen with or without login option. Users can select their residing state to get state-specific schemes. They can select if they want to look for 'Tribal Schemes' or 'Farmer Schemes'. On selecting one of the options, we further display categories of schemes they want to look for, like, Education, Health, Housing.

All the already available schemes will be displayed. New schemes will be added as and when they will be announced by the government. People who sign up on our app, can opt-in for user-specific notifications. Users who choose not to opt-in will get general notifications.

This design can be used to help farmers also. Similar to how we are displaying all the welfare schemes available to the tribal community, we can display all the schemes available to our farmers.

These days natural calamities have increased in number, mostly due to man-made problems. So, in such hard times our farmers can make the most use of the welfare schemes and reduce their expenditures.

IV. ADVANTAGES

- Tribal communities can find all welfare schemes of all states in one place.
- Farmers can find all the welfare schemes in one place.
- Very simple UI making it easy for new comers.

V. LIMITATIONS

- Redundancy of scheme related info.
- Assumption that user must be capable of using mobile application.
- As of today, the application supports only 1 language – English.

VI. FUTURE ENHANCEMENTS

For people who can use and know to make use of smartphones and laptops, searching and getting to know about government schemes is not a big deal, but for people who know very little about technology are losing the benefits provided by various state governments and also central government. This Tribal welfare has great application as it can be used by

people who have little knowledge of using mobile phones, as they can understand simple applications.

The system can be easily modified to help other backward communities and BPL community. If users are facing difficulty to read or use the app, we can work with Text-to-speech system. We can implement AI to add schemes and eliminate admin's involvement.

In future it is highly possible to improve the application in terms of speed, efficiency, reliability, adding more options once it gets developed.

VII.   CONCLUSION

As per the study in Anantapuram District of Andhra Pradesh, nearly 42.92 per cent of the beneficiaries were not aware of the facilities available under the scheme. They have missed many opportunities and many benefits also their expenditure has increased significantly. Making most people aware of the all the available schemes will not only benefit them to make use of the schemes, it also will reduce their expenditure they make on their health, housing and agricultural expenses.

This brings an awareness among people who are missing all the benefits and amenities that are included in the government schemes. Also, this project not only helps for us to improve our programming skills but, for the society also. We really hope this project helps many tribes and also people who live in rural areas by availing all the benefits provided by the government and get benefitted.

VIII. REFERENCES

[1]. https://nvshq.org/article/sarkari-yojana/.
[2]. https://rural.nic.in/more/mord-mobile-apps
[3]. https://www.newindianexpress.com/cities/kochi/
     2018/jul/18/an-app-that-can-update-you-about-
     all- government-schemes-1844695.html.

Life Saviour Detector

Ms. Yogitha[1], Chinmayi P Anvekar[1], M Gopinath[1], Jerin Jacob[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Outer Ring Rd., near Marathalli, Kadubessanahalli, Bengaluru – 560103, Karnataka, India

ABSTRACT

In a growing and developing country like India where population is increasing on daily basis, also results in congested roadways because of vehicles, animals, roadside shops. Due to which, accidents are at an all-time high nowadays. A survey states that India is a motorcycle dominated country which occupies about 65 percent of roadway vehicles used in India. In metropolitan cities like Bangalore accidents and its fatality rate is controlled due to educated population and attentive traffic agency and quick response from the health officials but recent statics show that despite measures being taken, accident rate is not decreasing even in metropolitan cities due to citizens negligence and carelessness. Under the new government, where a developing country like India is taking steps towards digitization, most of the traffic lights at crossroads are accompanied by cameras. Keeping the digital India in mind, this project is ambitiously aiming to develop a surveillance-based software which would detect if a particular rider is wearing a helmet. As from a bike rider point of view, the only and must security measure he/she could take is if he/she is wearing a helmet, wearing a helmet drastically reduces the fatality percent in an unfortunate event of an accident.

This project uses modern machine learning algorithms and latest R&D packages which uses various computer science fields like deep leaning, neural networks, and computer vision. This project can be used by various government agencies like traffic police etc. the government can use the software to check if a particular citizen is wearing a helmet and if not, the software can be developed to such a way which integrates the cameras all over the city creating a web which can be used to track the movements of the offender, depending on which the officials can take charge as mentioned in their jurisdiction.

Keywords— Helmet Detection, Object Detection, Deep Learning, Machine Learning, YOLO, Convolutional Neural Networks, Computer Vision, Training Data.

I. INTRODUCTION

The Life Saviour Detector, detecting of two-wheeler riders wearing a helmet or not and detecting helmetless two-wheeler license plate using Deep Learning based object detection algorithms. With varied deep learning algorithms, available we are moving forward with YOLO Real-time Object Detection algorithm. The leveraged real time object detection using YOLO algorithm is an algorithm based on regression, instead of selecting the interesting part of an Image, it predicts classes and bounding boxes for the whole image in one run of the Algorithm.

YOLO is built in such way that it approaches over the image/video frame at once and a single convolutional neural network is deployed over the image/video and resulting in formation of bounding boxes for the classes with the label and its confidence and here we have our classes as the helmet, motorbike, license plate and non-helmet rider and the result that we aim is to detect the license plate of helmetless motorbike rides. Computer vision using CNN (convolutional neural network) enables a computer to identify or detect process /object in image or video as humans do. Due to advances in AI and various innovations in deep learning and neural networks, this field has taken leaps in past years and has surpassed humans in some tasks of object detection and labelling.

A. Problem Definition

According to the Ministry of road transport highways, motorcycle accidents have been increasing throughout the years and when compared many lives were saved as they wore helmets and it is helping in preventing fatalities, So our main objective of this project is to build a model with the help of Deep learning and Object detection that can detect rider without helmet and raise a challan for the violation.

B. Objectives

This project aims at developing a helmet detection system that will be helpful for the system considering the current system where we only have a manual check point where traffic police raise a fine for a rider if he is found not wearing a helmet, but with the help of this system the defaulter's license plate will be captured and they will have to pay the respective fine. This system with the help of technologies like Deep Learning, Object Detection and the other required frameworks will be able to detect the riders without helmet.

C. Scope Of the Project

Two-Wheeler is a very popular mode of transportation in almost every country. However, there is a high risk involved because of less protection. To reduce the involved risk, it is highly desirable for bike riders to use helmets for their safety. Observing the usefulness of helmet, Governments have made it a punishable offense to ride a bike without helmet and have adopted manual strategies to catch the violators. The existing video surveillance-based methods are passive and require significant human assistance. Therefore, it is necessary to develop a system for automatic detection of two-wheeler riders if they are wearing a helmet or not. Therefore, a custom object detection model is created using YOLO Realtime Object Detection algorithm.

II. TECHNOLOGY

A. MACHINE LEARNING

Machine learning is a type of artificial intelligence that enables applications of software to become more precise in predicting results without being specifically programmed to do so. Various statistical and mathematical methods are used to measure the efficiency of ML models and algorithms. The trained model can be used after the conclusion of the learning process to identify, predict, or cluster new research data using the knowledge gained during the training process.

B. DEEP LEARNING FOR COMPUTER VISION

Computer vision is an area that involves making " see 'a machine. Instead of the human eye, this device uses a camera and computer to locate, monitor and quantify targets for further processing of images. Example of image processing includes:

- Normalizing photometric properties of the image, such as brightness or colour.
- Cropping the bounds of the image, such as centring an object in a photograph.
- Removing digital noise from an image, such as digital artifacts from low light levels.

With the development of computer vision, such technology has been widely used in the field of agricultural automation and plays a key role in its development. Deep learning in computer vision has made rapid progress over a short period. Some of the applications where deep learning is used in computer vision include face recognition systems, self-driving cars, etc. Convolutional neural networks, also known as convnets, a type of deep learning model universally used in computer vision applications. So, the objective of CNN is to perform two tasks: first is feature extraction and second is aggregating all the extracted features and making a prediction based on it.

C. YOLO v3 (YOU ONLY LOOK ONCE)

YOLO, another way to deal with object recognition. Earlier work on article recognition repurposes classifiers to perform discovery. All things considered, we outline object location as a relapse issue to spatially isolated jumping boxes and related class probabilities. A solitary neural organization predicts jumping boxes and class probabilities straightforwardly from full pictures in a single assessment. Since the entire location pipeline is a solitary organization, it tends to be streamlined start to finish straightforwardly on discovery execution.

Our bound together engineering is very quick. Our base YOLO model cycles pictures continuously at 45 edges for every second. A more modest variant of the organization, Fast YOLO, measures a shocking 155 casings for every second while as yet accomplishing twofold the guide of other constant finders. Contrasted with best-in-class location frameworks, YOLO makes more restriction mistakes yet is far more averse to foresee bogus recognitions where nothing exists. At last, YOLO learns general portrayals of items. It beats all other location techniques, including DPM and R-CNN, by a wide edge while summing up from common pictures to work of art on both the Picasso Dataset and the People-Art Dataset."

The entirety of the past item discovery calculations uses districts to limit the article inside the picture. The organization doesn't take a gander at the total picture. All things considered, portions of the picture which have high probabilities of containing the article. YOLO or You Only Look Once is an article recognition calculation vastly different from the district-based calculations seen previously. In YOLO a solitary convolutional network predicts the jumping boxes and the class probabilities for these crates.

III. SYSTEM ARCHITECTURE

The work has been developed using YOLOv3. First, we apply object detection algorithm YOLOv3 to obtain the two-wheeler riders. The bounding boxes obtained contain all the objects detected in the image belonging to dataset and it filters only the classes of persons and large vehicles. These bounding boxes are then cropped from the image and is trained to recognize the helmets from non-helmets.

- Data Source – open-source dataset v6, Datasets from repositories
- Image Processing – ML Model consisting YOLO v3 using Darknet, Convolutional Neutral Networks.

Results – Detection and Capture of default helmetless rider with license plate .

IV. PROPOSED SYSTEM

YOLO is state of the art object detection model which is fast and accurate especially when it comes to real time processing.

The system will take an image or video frame as input. The image will be forward propagation pass through the deep neural network as yolo takes the entire image in a single instance to make predictions.

Each image has a shape (608,608,3) as the framework is trained to run on 608x608 images. Further the framework divides input images into SxS grid, for each grid cell, consider a label y holding eight values as dimensional vector when we consideration person riding a bike, helmet and License plate as our 3 classes.

- pc denoting the probability of class/object present in the grid or not
- bx, by, bh, bw specify the bounding box, bx, by are the centroid
- bh, bw as the ratio of respective height and width of bounding box to the height and width of the corresponding grid cell.
- c means the class, e.g. (c=0 is Person with bike, c=1 is helmet)

Consider to expand c to a 3-dimensional vector, each bounding box is then represented by 8 numbers. The YOLO architecture considering 5 anchor boxes the output here will be processed in the fashion given below-:

Image (m, 608, 608, 3) passed to Deep CNN then Encoding (m, 19, 19, 5, 8).

For understanding when we will flatten the last two last dimensions of the shape (19, 19, 5, 8) encoding then the output of the Deep CNN is (19, 19, 40).

The class score calculation is done by $score\_c_i = p\_c *$ c_i Which means that the probability of presence of an object is p_c times the probability that the object is a certain class c_i.

After the calculation and prediction of score, box and class, the next step is the bounding box formation around the classes detected.

In each cell it can give 5 bonding boxes and the model is capable of predicting total of 19x19x5 = 1805 boxes with a single instance at the image. But for better results we required algorithm output with high accuracy, to achieve the we will be using NMS (Non-Max Suppression) to remove bounding boxes with less score or low confidence due to uncertainly of detecting a class or object.

Also overcoming the multiple bounding boxes overlapping for detecting the same class and get a single box in such case.

Solutions can be filtering the threshold on class scores and mainly when talking about NMS which uses concept of IOU i.e. Intersection Over Union for getting single bounding box for overlapping boxes created during the detection of the same object.

iou >= iou_threshold

Non-Max Suppression, in this is process involves taking high score boxes and applying the formula above and continuing this until we are left with the high score boxes and the best box found at the end. This is the entire proposed system and the working of the yolo for our project life saviour detector.

V. REQUIREMENT ANALYSIS

In software development, functional requirements define the function or part of the software system. A function is a collection of inputs, outputs, and actions. Computing, technical information, data processing and processing, and other basic characteristics that determine the purpose of the system design may be functional requirements.

The environment includes users, administrators, and other external systems that interact with them.

The testing image is given to the system and the system resizes the given image to threshold size. The system takes the resized image and compares with the

trained weights. After comparing, objects are detected and then bounding boxes are created around it.

Hardware and Software Requirements

1) Hardware Requirements:
- Processor: i5 Processor
- RAM: 8GB
- Hard Disk: 10 GB
- Any desktop / Laptop system with above level configuration or higher level.
- GPU: NVIDIA 1080 or NVIDIA Titan

2) Software Requirements:
- Operating system: Windows 10
- Language : Python
- IDE: Jupyter Notebook, Python IDE, Anaconda, Google Colab
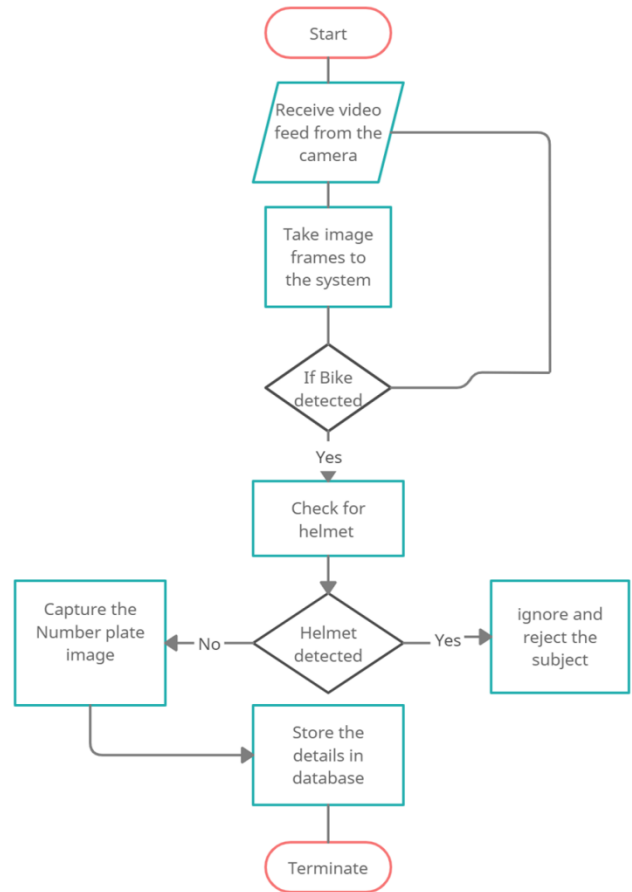
VI. DESIGN

A. Design diagrams
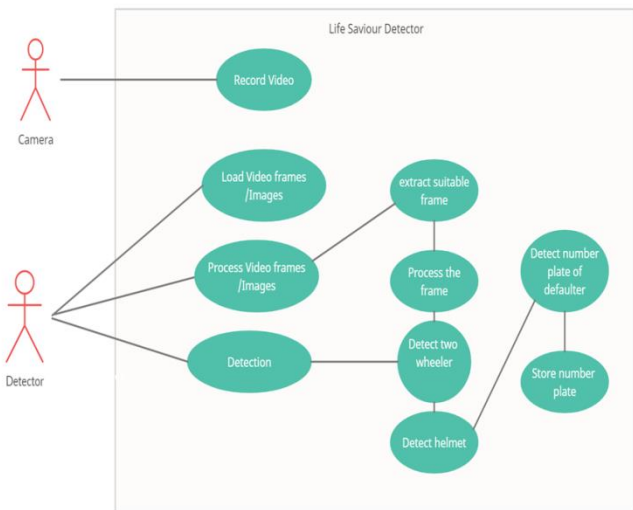
Fig.6 Use Case Diagram

Fig.7 Activity Diagram
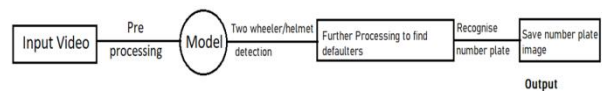
Fig.8 DFD Level 0
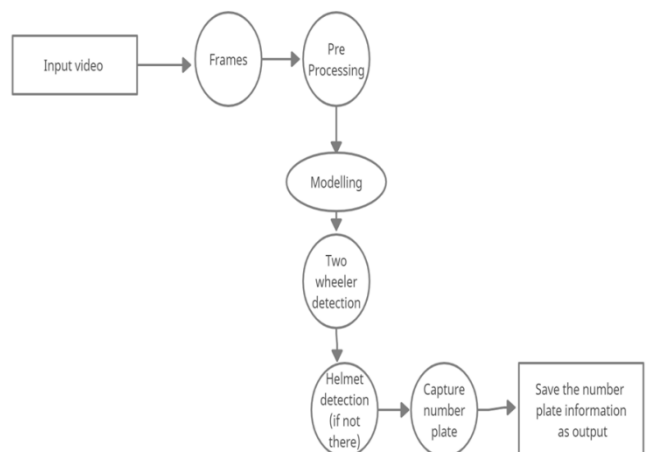
Fig.9 DFD Level 1

VII. CONCLUSION

Through this project we want to develop a system which can help in bringing awareness among people who use two-wheelers without wearing helmets. We bring forward a framework for real-time detection of traffic rule defaulters who ride bike without using helmet. By using various technologies like Deep Learning, Objection detection we aim in bringing good results and increase the efficiency of the system and help the traffic control department in finding out the defaulters easily rather than depending on the current system where cops find the defaulters and raise a fine, in this current system it is not efficient because not all the defaulters are found guilty. So, by creating such a system we feel people will be more careful and will not be negligent towards wearing helmets and by this we can avoid many accidents and save many lives.

VIII. ACKNOWLEDGMENT

IX. REFERENCES

[1]. https://www.sciencedirect.com/science/article/pii/S2405959519304187

[2]. https://link.springer.com/chapter/10.1007%2F978-3-319-590721_10

[3]. https://ieeexplore.ieee.org/document/8310092

[4]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm

[5]. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[6]. https://medium.com/@romanovacca/detecting-scooter-drivers-using-a-custom-object-detection-model-based-on-yolov3-501ec3d1a018.

[7]. https://iopscience.iop.org/article/10.1088/1742-6596/1821/1/012049

[8]. Lini shine, Siji C.V -Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN

[9]. C. Vishnu, Dinesh Singh, C. Krishna Mohan and Sobhan Babu - Detection of Motorcyclists without Helmet in Videos using Convolutional Neural Network

[10]. Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks,"

[11]. Yange Li, Han Wei, Zheng Han, Jianling Huang and Weidong Wang - Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks

[12]. Kelson R. T. ,Aires,Rodrigo ,Veras,Romuere Silva -Helmet Detection on Motorcyclists Using Image Descriptors and Classifiers

[13]. Prajwal M. J., Tejas K. B., Varshad V., Shashidhar R -A Review on Helmet Detection by using Image Processing and Convolutional Neural Networks

Drowsiness Detector

Aayush Soni[1], Jehan K E[1], Bishal Soni[1], Sivabalan[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore,
Karnataka, India

ABSTRACT

Today drowsy driving is a serious problem that leads to thousands of accidents each year. Nowadays, accidents involving vehicles are extremely commonplace and cause lot of damages to people monetarily and physically. Drowsiness is one of the factors for collisions. In India, no monitoring device is used to measure the drowsiness of the driver. Some kind of monitoring system like driver fatigue monitoring, real time vision based on driver state monitoring system, seeing driver assisting system and user center drowsiness driver detection system are implemented in foreign countries. With the help of drowsiness detector, we aim to track the movement of eyes when the person feels tired. If the eyes of the driver are shut for an ample duration, the driver will be alerted through an alarm and some tips will be shown to the driver to get rid of their drowsiness.

Keywords— Drowsiness, fatigue, eyes, alarm.

I. INTRODUCTION

Today drowsy driving is a serious problem that leads to thousands of accidents each year. Motor vehicle collisions lead to significant death and disability as well as significant financial cost to both security and individuals due to the driver impairments. Drowsiness is one of the factors for collisions. In India, no monitoring device is used to measure the drowsiness of driver. Some kind of systems like driver fatigue monitor, real time vision based on driver state monitoring system, seeing driver assisting system, user center drowsiness driver detection and working system are implemented in foreign countries. According to a lot of studies driver drowsiness is leads to a large portion of accidents on a daily basis all over the. Hence it is a large problem that has to be addressed. The aim of this is to detect whether a person is drowsy or not by detecting the users' eyes through smartphone cameras from the driver. The data collected is then used to calculate an eye closure percentage and compared with a predefined calculated threshold value to check and ensure whether the driver is drowsy or not.

II. RELATED WORK

Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks: A Survey

Real time drowsiness detection using eye blink monitoring

According to analysis reports on road accidents of recent years, it's renowned that the main cause of road accidents resulting in deaths, severe injuries and monetary losses, is due to a drowsy or a sleepy driver. Drowsy state may be caused by

lack of sleep, medication, drugs or driving continuously for long time period. An increased rate of roadside accidents caused due to drowsiness during driving indicates a need of a system that detects such a state of a driver and alerts him prior to the occurrence of any accident. During the recent years, many researchers have shown interest in drowsiness detection. Their approaches basically monitor either physiological or behavioural characteristics related to the driver or the measures related to the vehicle being used. A literature survey summarizing some of the recent techniques proposed in this area is provided. To deal with this problem we propose an eye blink monitoring algorithm that uses eye feature points to determine the open or closed state of the eye and activate an alarm if the driver is drowsy. Detailed experimental findings are also presented to highlight the strengths and weaknesses of our technique. An accuracy of 94% has been recorded for the proposed methodology.

A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection

Drowsiness can put lives of many drivers and workers in danger. It is important to design practical and easy-to-deploy real-world systems to detect the onset of drowsiness. In this paper, we address early drowsiness detection, which can provide early alerts and offer subjects ample time to react. We present a large and public real-life dataset of 60 subjects, with video segments labelled as alert, low vigilant, or drowsy. This dataset consists of around 30 hours of video, with contents ranging from subtle signs of drowsiness to more obvious ones. We also benchmark a temporal model for our dataset, which has low computational and storage demands. The core of our proposed method is a Hierarchical Multiscale Long Short-Term Memory (HM-LSTM) network, that is fed by detected blink features in sequence. Our experiments demonstrate the relationship between the sequential blink features and drowsiness. In the experimental results, our baseline method produces higher accuracy than human judgment.

III. ALGORITHM

Convolution Neural Network method uses layers of spatial convolutions that are well suited for images, which exhibit strong spatial convolutions. The blinking of only one eye, either right or left, is detected so that the memory for detecting both the eyes is saved. We only need one eye blinking information because we blink both the eyes together. The threshold value is a preset value which indicates an eye closure point beyond which the eye of the average user can be detected as closed. This value is predetermined through several conducted studies to

be around the 70% mark or in this case around the value of 0.75.

Permissions are extremely essential as they allow the user to know what privacy services the application is using and allows the users to approve or deny access. PackageManager is used in order to obtain these permissions from the user in Android Studio. The user is shown a popup to grant the permission when the application is started for the first time.

$$PERCLOS = \frac{Number\ frames\ of\ "closed"\ eye}{3\ min\ interval\ of\ all\ frames - blinking\ time}$$

In this system the recommended PERCLOS alarm threshold of 0.15 is used as the highest level of drowsiness. As the following table shows the drowsiness level which is based on the PERCLOS thresholds (% of eyes closure over the 3 minutes interval).

Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5
S≤3.75%	3.75%<S≤7.5%	7.5%<S≤11.25%	11.25%<S≤15%	15%<S
Low drowsiness	Low drowsiness	Moderate Drowsiness	Moderate Drowsiness	Severe Drowsiness

For each eye, there are six points distributed around to locate the position of eye. The distribution of eyes landmarks has significant differences between open and closed state. Eye Aspect Ratio was application to record the blink frequency. EAR can be computed according to the position of eyes landmarks by:

$$EAR = \frac{\|P_2 - P_6\| + \|P_3 - P_5\|}{2\|P_1 - P_4\|}$$
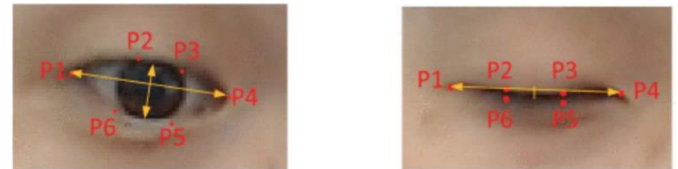
The above plotted landmark P points are used to check if the users eyes are open or not by comparing with the predetermined threshold value that is set in order to check if the user is drowsy or not.

(a) f-5874   (b) f-8559   (c) f-27509   (d) f-27647

(e) f-30102   (f) f-32249   (g) f-39407   (h) f-39944
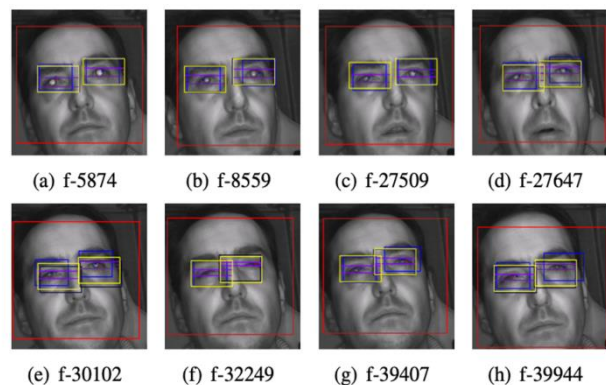
IV. METHODOLOGY

To detect and track eye images with complex background, distinctive features of the user eye are used. Generally, an eye-tracking and detection system can be divided into four steps: Face detection, eye region detection, pupil detection and eye tracking. To find the position of the pupil, first, the face region must be separated from the rest of the image using a mixture of Gaussian, this will cause the images background to be non-effective in our next steps. We used the horizontal projection obtained from the face region, to separate a region containing eyes and eyebrow. This will result in decreasing the computational complexity and ignoring some factors such as bread. Finally, in proposed method points with the highest values are selected as the eye candidates. The eye region is well detected among these points. Colour entropy in the eye region is used to eliminate the irrelevant candidates. In the next step, we perform eye tracking. In the proposed method, eye detection and tracking are applied on testing sets, gathered from different images of face data with complex backgrounds. Experiments indicate a correct detection rate of 94.9%, which is indicative of the method's superiority and high robustness.

Here the process conducted in various modules:

A. Face and Eye detection.
B. Eye Image Filtering.
C. Eye closure evaluation.
D. Drowsiness parameter estimation.

As mentioned, to a certain extent, the state of eyes indicates whether the driver is drowsy or not. Because there are significant differences about time of eyes closed between awake and drowsy. A method of ellipse fitting was proposed to describe the shape of pupil. The method segments the pupil with traditional image process firstly. Then, an ellipse is fitted with the white pixels, which represent the shape of eyes. Lastly, the ratio of the major and minor axes of the ellipse was used to evaluate the eyes state.

In case of android, we have come across vision. Face which consists of multiple sub libraries such as Face, Face Detector, Landmark. A human face detected in an image or video. It is important to note that all fields described here are with regards to the image that the detector has processed. Many live apps that process images directly from the camera show the user a mirrored display of the actual image. All coordinate values are reported as absolute image coordinates. That is, image position (0, 0) represents the upper-left corner of the image. From this Face library we have multiple public methods, from which we can get different kinds of values, public methods are given as:

- **public list<Contour>getContours()** - Returns a list of contours (eyes, nose, etc.) found on the face. A contour detector must be specified via FaceDetector.Builder.setLandmarkType(int) to detect contours. The contour detector may not find all possible contour on any given face which gives a list of landmarks found on the face.
- **public float getEularY()** - Returns the rotation of the face about the vertical axis of the image. Positive euler y is when the face turns toward the right side of the of the image that is being processed which returns the rotation of the face about the vertical axis of the image.
- **public float getEularZ()** - Returns the rotation of the face about the vertical axis of the image. Positive euler y is when the face turns toward the right side of the of the image that is being processed which returns the rotation of the face about the vertical axis of the image.
- **public float getHeight()** - Returns the height of the face region in pixels. This is a rough estimate that is likely to be slightly larger than the exact bounds of the face and therefore may include some background which returns the height of the face in pixels.
- **public int getId()** - Returns the face ID. This can be used to track a Face over multiple Frames.
- **public float getIsLeftEyeOpenProbability()** - Returns a value between 0.0 and 1.0 giving a probability that the face's left eye is open. which returns the probability for the face's left eye being open
- **public float getIsRightEyeOpenProbability()** - Returns a value between 0.0 and 1.0 giving a probability that the face's right eye is open which returns the probability for the face's right eye being open.

V. CONCLUSION

Drowsiness detection is a precautionary software that can aid in avoiding accidents that are due to drivers who are falling asleep while driving. It is a vehicle safety software which will aid in prevention of accidents caused because of the drivers getting drowsy. Drowsiness detection system which can help control the accidents which in return is going to help reduce the deaths of people due to road accidents. This system proposes a drowsiness detection software that works on image processing using machine learning algorithms. It is designed with embedded systems in mind. This is a system used in automotive vehicles. The system is capable of detecting whether the user is drowsy or not in a very short duration of time. After the detection of abnormal behaviour, it is alerted to the driver through alarms.

VI. REFERENCES

[1]. Harry Zhang, Gerald J. Witt and Matthew R. Smith, "Drowsiness detection system and method" Publication of US7202792B2 : Coming of Age: 2007-04-10,Current Assignee: Aptive Technologies Ltd ,Application no. US10/291,913.

[2]. H.Ueno, M. Kaneda and M.Tsukino, "Development of drowsiness detection system" in Proc. Proceedings of VNIS'94 - 1994 Vehicle Navigation and Information Systems Conference : Coming of Age: 31 Aug.-2 Sept. 1994, IEEE, DOI: 10.1109/VNIS.1994.396873.

[3]. "Overview of research on driver drowsiness definition and driver drowsiness detection", National Highway Traffic Safety Administration, 1995. [Online]. Available: https://www.safetylit.org/citations/index.php?fuse action=citations.viewdetails &citationIds[]=citjournalarticle_245681_38

[4]. Richard Frederic Kaplan and Kenneth Alan loparo ," Alertness and drowsiness detection and tracking system", Publication of US5813993A,Coming of Age: 1998-09-29,Current Assignee: CONSOLIDATED RESEARCH OF RICHMOND, INC, Application no. US08/628,474.

FUTURES - The Career Predictor

S.P.L Santoshi[1], Prathiksha S.P[1], Chandana Menon[1], Mr.Kiran Kumar K[2]

[1]B.Tech Student, Department of Computer Science, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Assistant Professor, Department of Computer Science, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Nowadays, many students are confused about their future. Choosing the right career is the toughest decision. Students are going through hard times to pursue the correct career.so many people suggest different career but only they know what they really want to do.so the application which is being developed will be helping the ones who are confused about their career path.by using machine learning algorithms and other prediction models. this application is going to take some inputs about the person who is seeking a job related to academic, personal interests etc and predicts a suitable career to them based on their data given.

Keywords: Career, Prediction, Naïve bayes, Machine Learning, Data mining

I. INTRODUCTION

Career can be known as a job or occupation of any person in which one will be learning throughout the process , improving their knowledge and skill set where they will be earning money also.one should be planned and organised to get a deserving career in this competitive technical world. So everyone who is so concerned about getting a right career must be evaluating their skills, improving them according to the updating tech world. Their skill set will improve by these processes.

The companies evaluate the job seeking people on different parameters like their personal interests, academic details, projects they have done, internships they have gone through, workshops they have attended, exams they have taken etc. and give them a set of rounds of examination, intake only those who have cleared them with minimum cut-off percentage.

The application which is being developed will help the people who are willing to choose the career path. The application suggests the right career to a person based on their inputs and data given to the system. It suggests few job roles which suit his profile.

This application is developed using naïve bayes theorem. It is a learning algorithm which is based on Bayes' rule. To divide the data into independent classes and calculate the distribution of each class this bayes theorem is used. The main reason to use this to develop our project is that we will be using a large set of data and naïve bayes theorem is easier to train and test the data

II. SYSTEM ANALYSIS

A. Existing System

There are so many previously developed systems for job seeking. All these existing systems like AMCAT,

COCUBES provide job recommendations based on aptitude , logical reasoning ,english comprehension etc people who cleared these exams will be undergoing set of interview conducted by the companies and some other sites like LINKEDIN, GLASSDOOR etc just lets the user to search the jobs and apply to them and suggests some jobs based on their previous experience and technical skills in their profile. but the proposed system is quite different from all these which predicts the job based on the person's interest,academic details etc.

B. Proposed System

In today's world students are very confused about choosing the right career path.so the proposed system will help the students or anyone who is seeking a better career path in choosing the career by predicting the career growth based on given data.

The user who seeks the career prediction have to login to the   application, if they are a new user,they have to create a new account by registering where they give all basic details required to login.once the user is logged in,they will be provided with options to fill details about their academic details, project information,workshops or internships attended etc. and then they will be having n option to upload the documents related to the same which will be validated later by admin and can also take the career prediction. This system predicts the job according to the data given.

If the user is admin, he will be logged in using particular requirements.the job of admin would be to analyse the data on information gathered together, manage the users like to accept or reject their profile after validating the documents.

Naïve Bayes theorem is being used in the implementation of the system as it is highly accurate when applied to big data sets as it requires training and testing dataset.

III. DIAGRAMS
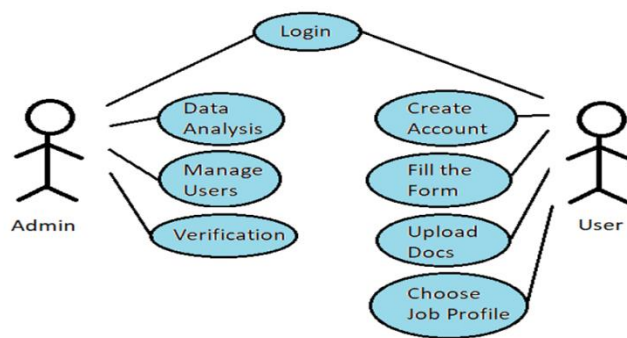
Fig.1. Use Case
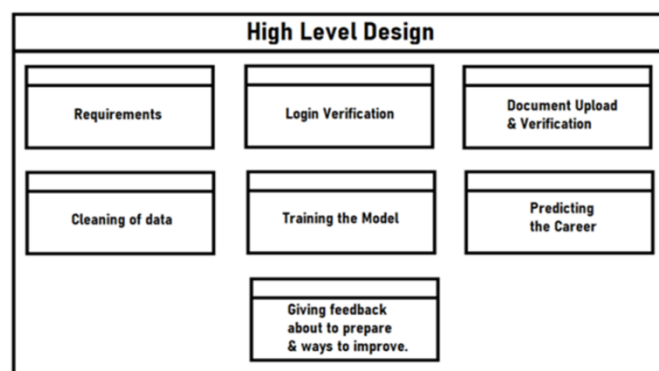
Fig.2. High Level Design

IV. SYSTEM IMPLEMENTATION

A. Collection of Data

The data consist of Academic percentage, percentage in Operating Systems, percentage in Algorithms, Percentage in Programming Concepts, Percentage in Software Engineering, Percentage in Computer Networks, Percentage in Electronics Subjects, Percentage in Computer Architecture, Percentage in Mathematics, Percentage in Communication skills, Hours working per day, Logical quotient rating, hackathons, coding skills rating, public speaking points, can work long before the system, self-learning capability, Extra-courses did, certifications, workshops, talent tests taken, Olympiads, reading and writing skills, memory capability score, Interested subjects, interested career area ,Job/Higher Studies, Type of company want to settle in , Taken inputs from seniors or elders, interested in games, Interested Type of Books ,Salary Range Expected, In a Relationship, Gentle or Tuff behaviour, Management

or Technical, Salary/work, hard/smart worker, worked in teams ever , Introvert, Suggested Job Role.

B. Label Encoding

We need to transform the columns which store text data to a unique number which represents them. Label encoding helps in converting text to significant numbers. We are using Label Encoder for the conversion.

C. Training and Testing

Once the data is cleaned and converted to the form required we use train_test_split to split the data and train it. Test data size is 33%.
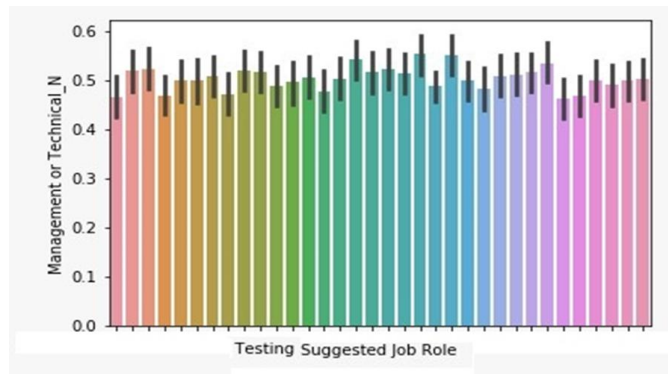
Fig .3.Management /Technical Vs Testing Suggested Job Role
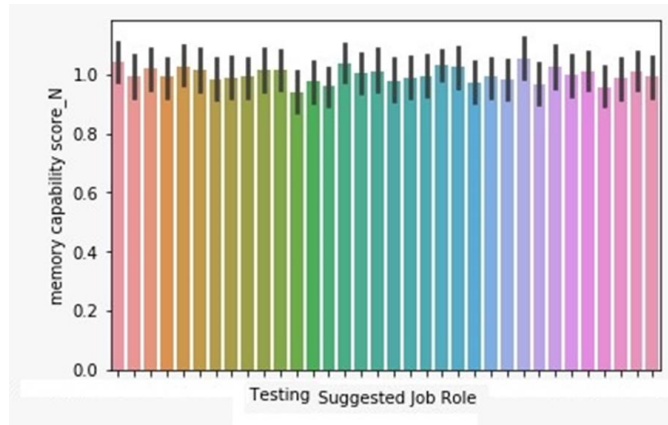
Fig .4.Memory Capability Score Vs Testing Suggested Job Role
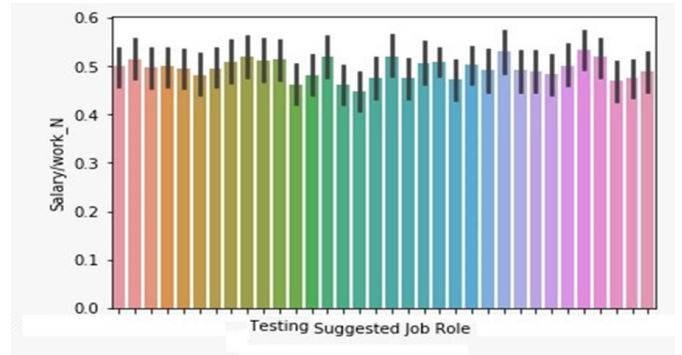
Fig .5.Salary/work Vs Testing Suggested Job Role

D. Algorithm Implementation

Naïve Bayes algorithm is used for predicting as it is more efficient in dealing with huge data. Sklearn provides libraries for dealing with machine learning algorithm.GausianNB is imported from sklearn.naive_bayes.Classifeier is an object of GaussianNB. Once the text data is converted to numbers with the help of label encoding the text and respective data are stored in dictionaries which can be used for converting the newly obtained data's text to there respective numbers without calling label encoder again.The model is fitted by calling fit function whose parameters are x_train and y _train.X_train contains the details which help in deciding the y_train values. Y_train contains the output i.e Job Role.Once the model is fitted we can predict the Job role by calling predict function by passing x_test as a parameter.
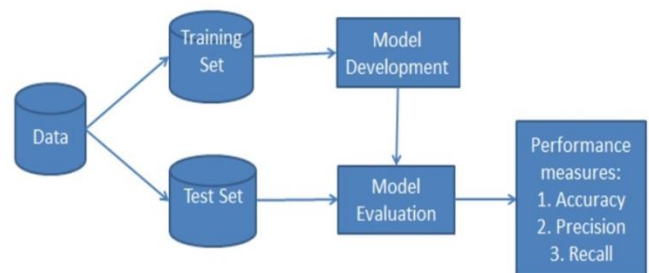
Fig.6. Naive Bayes Algorithm

E. Accuracy

Accuracy of the built model can be found by using accuary_score which is available in sklearn.metrics.y_test and predicted variables which

store the prediction are passed as a parameter to accuary_score.The accuracy can be increased by improving the data passed and there form.

F. Result

The data given by the user is evaluated in the backend using the algorithm built and displays the Suggested Job Role to the user in the front end.

V. CONCLUSION

A prototype of a career prediction system is developed using Naive Bayes Algorithm. The system extracts hidden knowledge from the historical Career prediction database. The models are first trained and are validated against a test dataset. Prediction, Identification, Classification, Optimization that are considered to be the goals of data mining are defined based on the data exploration. The goals of data mining are evaluated against the trained model. Engineering students of CS/IT can use this Career Prediction system to evaluate their skills, talent, interests and find the career path and career that is best fit for them.

VI. FUTURE SCOPE

A more powerful application can be developed where student parameters are taken directly by evaluating students through various evaluations and examinations. Systems can be made more reliable to use by developing Technical, analytical, logical, memory based, psychometry and general awareness, interests and skill-based tests through which parameters are collected so that the results can be certainly accurate.

We all know Education is evolving day-by-day with the help of the internet that has numerous videos, study materials and tutorials. This Application will also provide numerous videos, study materials and tutorials that can be used by the students to build

their understanding in the field of their interest. There are many other fields emerging in the Engineering Field. We will expand our scope by predicting the career from other fields such as Electrical Engineering, Electronics Engineering, Mechanical Engineering etc including CS/IT.

VII. REFERENCES

[1]. Student Career Prediction Using Advanced Machine Learning Techniques -K. Sripath Roy , K.Roopkanth , V.Uday Teja , V.Bhavana, J.Priyanka .International Journal of Engineering & Technology, 7 (2.20) (2018) 26-29

[2]. Career Guidance System using Machine Learning For Engineering Students -Ankush Daharwal, Prof. Sandeep Gore, Aishwarya Bhagwat, Shraddh, Dethe,Sunny Chavan.International Research Journal of Engineering and Technology (IRJET) ,Volume: 07 Issue: 06 | June 2020

[3]. Student job Prediction And Analysis - Ditika Bhanushali, Seher Khan, Mohommad Madhia, Shoumik Majumdar.International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 CertifiedVol. 7, Issue 3, March 2018

Intelligent Block Energy

Ruman Ahmed Shaikh[1], Sujay Hazra[1], Nikhil Sumesh Babu Nambiar[1], Uma.N[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Outer Ring Rd., near Marathalli, Kadubessanahalli, Bengaluru , Karnataka, India

ABSTRACT

The project proposes a framework on the basis of blockchain and deep learning for Smart Grids. The framework is built on the basis of 2 schemes, i.e., blockchain-based scheme and deep-learning based scheme. The blockchain-based scheme consists of four phases; setup phase, agreement phase, creating a block phase and consensus-making phase. It constitutes a dependable peer-to-peer energy system which follows the practical Byzantine fault tolerance algorithm reaching throughput. Prevention of attacks on the smart grid requires the framework to generate blocks with the help of short signatures and hash functions. In order to prevent fraudulent transactions and network attacks in the network, the proposed scheme developed on the basis of deep-learning, i.e., an intrusion detection system (IDS), that utilizes neural networks, has been introduced.

Keywords — Blockchain, Cryptocurrency, Wallet, Public Key Private Key, Deep Learning, Recurrent Neural Network

I. INTRODUCTION

As climate change and its effects are getting worse and more devastating, the need for a smart grid system powered by renewable energy sources increases more and more. A smart grid is an interconnected network of internet of things that can share energy with one another. This means an entity with excess energy can sell energy while an entity with insufficient energy can buy energy, which makes the renewable energy source much more reliable. The problem with using an existing tech stack to build a smart grid is that it either relies on a grid-connected (centralized) or off-grid (isolated). The grid-connected or centralized solution relies on a central server which is very likely to go offline during a disaster and requires the users to trust a central entity (government or company). On the other hand, an off-grid or an isolated solution means the user is completely on its own. If the system loses power and exhausts its backup, then there is no way for it to get more, leaving its user in literal darkness.

In order to overcome these problems, we propose a blockchain based system whose security is reinforced with deep learning based cyber security measures like Intrusion Detection System (IDS). The ability of blockchain to establish consensus in a trust-less network will allow the system to establish value without the control of any single entity. This means the government or any company will not be able to set a price of the energy as they are peers in the network, if they are even present in the first place. As blockchain is relatively new technology. It has a lot of cybersecurity flaws; these flaws are enforced using AI based network security software.

A. Problem Definition

In this project, framework for the smart grids are developed using deep learning and blockchain. The management of electricity distribution, through the technologies available, can be settled by generating the smart grids using a set of controllers and computers with automation, and communication protocols, connected via the Internet.

A solar panels' network in the smart grid could create energy storage to merge the city's overall energy infrastructure. One important matter that we need to deal with during the smart grid's development is mainly in ensuring both privacy and security and not in the physical support, which is a main issue to the cyber security research community. There could be an adversary which launches interior and exterior attacks (e.g., false data injection and DNS attacks) which would disturb the smart grid's operation. One of the examples would include modification operations on the electricity data via eavesdropping attack or man in the middle attack and distributed denial of service DNS attacks on the communication of network protocols (i.e., TCP/IP, HTTP, UDP).

B. Objectives

1) The basic objective of the project is to create a smart grid which produces secure data transactions between the two parties. In order to produce secure data transactions between two parties, the concept of blockchain is being introduced.

2) Managers have to use technology of blockchain at later stages for problem solving and create new opportunities across the industries.

3) Even though blockchain is so promising, it is still vulnerable to attacks when it comes to data privacy. These issues can be dealt with technologies such as intrusion detection systems (IDS)and techniques of machine learning.

4) Therefore, in the project, energy framework is made using deep learning and blockchain for protecting the grid from cyber-attacks. The framework facilitates the excess energy exchange among neighboring nodes.

5) The scheme based on deep-learning is proposed using a artificial neural network (ANN) used in the detection of attacks on the network and harmful transactions.

6) This is the study that brings a combination of blockchain technology with deep learning into one architecturally secure framework for this smart grid.

C. Scope of the Project

The scope of this project is the distribution of electricity energy. The project provides a smart and secure way to distribute and trade energy in a decentralized way. The software is going to be packaged into a hardware unit, the unit has to be connected to bi-directional energy grid and to the internet. It will consist of a powerful Graphical Processing Unit (GPU) which can be again shared between multiple users to minimize cost. The computer is basically a computer with a dedicated GPU.

The software is going to be running on this system, which will bring in the blockchain based distribution and trading scheme, and a system based on deep learning for security. The main user of this system will be individuals who have access to a bi-directional energy grid, who have the cost incentive to afford the hardware and who have solar panels or other renewable source of energy. Once more and more of these systems gets added to the network, the cost per watt will fall. This will enable more people to invest as in the long run it will pay for that investment. A study suggests that the richest 1% causes more carbon emission than the poorest 50%. Therefore, by making the richest use this system a lot of carbon emission can be cut down. This can be only achieved if the government takes climate change and impose large taxes on the rich.

The virus causing respiratory ailment i.e., the coronavirus started in 20l9 have killed more than 1.5M+ deaths as of 5th December 2020. Whereas the World Health Organization (WHO) estimates that 150,000 people are dying per year due to the effects of climate change. This is expected to rise 250,000 by 2050 which puts the total at 100M people. This is true if we don't do anything about it right now, because unlike a pandemic there is no vaccine. There are many points of no return in when it comes to climate change. And obviously, if we don't do anything about it then it's going to be an extension level event.

So, the main scope of this project is survival and not just money. The incentive is survival of the human race very literally. It will help slow down climate change by making renewable energy more stable and reliable by smartly managing its distribution. This project is very dependent on the availability of the underlying hardware as it is a software as a service (SAAS). The scope of this project depends on how well the renewable ecosystem is developed and how well aware the current government is aware about climate change.

II. LITERATURE SURVEY

A. Technology

1) Blockchain: The most important step in the software development cycle is nothing but the literature survey. Before the development of the tool, we need to necessarily determine the factor of time, economy and strength of the company. Once the things are satisfied, then the next step is to have a look on which operating system and language is optimal for development of the tool. A lot of external support is required by the programmers while the process starts. This support is obtained from the senior programmers, from books or from the Internet. Before the building of system, the above consideration should be considered for developing the proposed system. We have to do analysis the existing smart grid systems. Any centralized system always has a single point of failure (SPOF).

In order to eradicate this problem few solutions were introduced, including distributed computing and peer-to-peer networking. The most novel solution utilizing both the mentioned accepts is blockchain, which was invented by a person (or a group of people) using the name/alias Satoshi Nakamoto in the year of 2008 to serve as the technology behind bitcoin.

Blockchain is a decentralized distributed ledger. The blockchain is chained through cryptographic hash in a block which contains the hash of the previous block, transaction details, timestamp, etc. It can be used as a distributed database which in case of a conflict of data can come to a consensus using some clever algorithms like proof work, proof of stake, longest chain and 51%. Each of these consensus algorithms have its own advantages and disadvantages. The cryptographic hashing of data inside the blocks makes it immutable, making it tamper proof.
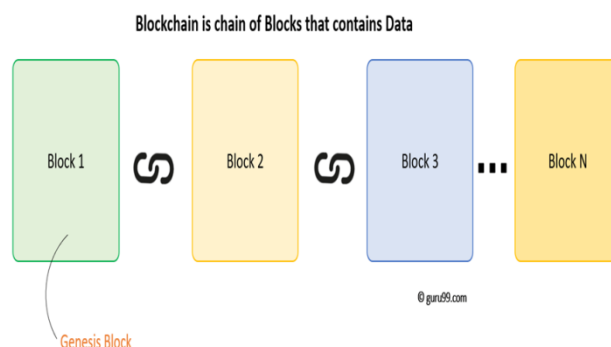
Fig.1 Chain of Blocks

2) Deep Learning: Deep learning is a blend of artificiality (AI) and machine learning (ML) that imitates the way humans gain certain types of knowledge. Deep learning is a major element of data science, It is extremely beneficial to data scientists who are tasked with collecting, analyzing and interpreting large amounts of data; deep learning makes this process faster and easier. At its simplest, deep learning could be thought of as a way for automate predictive analytics. While these machine learning algorithms are traditional and linear, there is stacking in hierarchy for the algorithms of deep learning to increase complexity and abstraction.

Neural are layers of nodes, much like the human brain is made up of neurons. Nodes within individual layers are connected to adjacent layers. The network is said to be deeper based on the number of layers it has. In an artificial neural network, signals travel between nodes and assign corresponding weights. A heavier weighted node will exert more effect on the next layer of nodes. The final layer compiles the weighted inputs to produce an output. Deep learning systems require powerful hardware because they have a large amount of data being processed and involves several complex mathematical calculations. Even with such advanced hardware, however, deep learning training computations can take weeks.

Deep learning systems require large amounts of data to return accurate results; accordingly, information is fed as huge data sets. When processing the data, artificial neural networks are able to classify data with the answers received from a series of binary true or false questions involving highly complex mathematical calculations. For example, a facial recognition program works by learning to detect and recognize edges and lines of faces, then more significant parts of the faces, and, finally, the overall representations of faces. Over time, the program trains itself, and the probability of correct answers increases. In this case, the facial recognition program will accurately identify faces with time.

3) Combination of Blockchain and Deep Leaning: Blockchain and machine learning, both are modern technologies that have materialized in the last decade. Machine learning provides the computer the ability to learn over time without being repeatedly programmed and without any human intervention. On the other hand, blockchain's primary function is making secured transactions between participants. Machine learning is a technology that relies on extensive quantities of data for model building and accurate prediction. This is where blockchain comes into play as the time taken can be considerably decreased by using blockchain technology. By using smart contracts in this case, data can be transferred directly and securely. For example, a machine learning model for self-driving cars would require hundreds of Terabytes of actual car driving data. Traditionally, using different trackers, all of the data like fuel consumption, driving speeds and breaks would be collected.

Then, it would be sent for processing where auditors would analyze the data to make it authentic and free of any discrepancies before sending it for processing to data scientists. However, smart contracts, by using digital signatures, could considerably improve the whole process. To certify the security of the data collected, using blockchain technology, smart contracts could be programmed to directly send data from the car driver to data scientists who will use the data for building machine learning models.

4) Flask: FIask is a lightweight WSGl web application framework. It is designed for swift and quick launches, with the ability of scaling up to applications that are complex. Flask offers suggestions, but when it comes to project layout, dependencies are not enforced. The developer chooses the tools and libraries which is needed. The community provides various extensions that make adding of new functionality easy.

Flask is considered more Pythonic than the Django web framework because in common situations the equivalent Flask web application is more explicit. Flask is also easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running.

B. Existing System

The current system for smart grid includes, grid-connected systems and off-grid systems.

In a grid-connected system, the user has a bi-directional line with the energy grid. This allows the user to transfer excess energy to the grid for reward or draw energy from the grid at a certain rate. The technical flaw of this system is SPOF which occurs mostly in natural disasters. In this scenario, if the user

doesn't have a better backup system, then the user loses power completely.

On the other hand, the off-grid system is completely isolated and thus, doesn't have any means of energy trading with anyone else. This means if the system cannon generates enough power or exhaust its backup power, then it's left in the dark.

A grid-connected system with battery backup can stay powered for a while, thus solving the SPOF problem for a certain time. But it still depends on a centralized hub for energy trading and therefore give up control on energy pricing. If a neighborhood can trade energy among themselves during natural disasters or otherwise, it will be much more efficient than relying on a centralized entity.

III. PRELIMINARIES

A. Proposed System

Our system is going to let users trade energy among themselves using cryptocurrency whose transaction are stored in a blockchain. This blockchain is synchronized among all the peers in the network by achieving consensus using the Satoshi Nakamoto consensus algorithm introduced in the blockchain white paper. This trading is done using the smart contracts to ensure fair trading.

1) Problems Solved by the proposed system are as follows:
- Keeps the advantages of using a blockchain
- By using smart contract, the delays and enounces of energy transmission like transmission rate, battery's discharge and charge rates, etc. are abstracted away.
- The use of blockchain make this grid very suitable for renewable energy sources like solar energy.

B. System Components Analysis

This section describes the conceptual and the implementation details of individual component in the system:

1) Cryptocurrency: A cryptocurrency is nothing but a digital medium of exchange. It has three main aspects in its technology i.e., securing blockchain, wallets and the concept of mining. Cryptocurrency does leverage the blockchain in order to allow a public database of transactions to be accessed by everyone. Since the transactions are accessible to the whole world, there are always possibilities of adversary attacks.

2) Digital Signature: Therefore, a phenomenon called cryptography is used by the cryptocurrencies in order to prevent these attacks and protect the blockchain. Cryptography is the process of using various algorithms to obscure the data available into hidden, or cryptic, messages. The main operation of cryptography, within a cryptocurrency, is to allow individuals to generate unique digital signatures.

Fig. 2 Digital Signature

Digital signatures are hand-written signatures with enhanced level of security which solves the problem of tampering. Therefore, each individual who wants to record their transaction into the blockchain has to stamp their data with a digital signature. The signature is based on cryptographic key, where one key is a public key and the other key is the private key. The key pair of the public and private keys are stored in objects called wallets (or) crypto wallets.

3) Crypto Wallet and Transaction Pool: Wallets are objects that store the key pair of the public and private key of and individual. It helps us understand (or) identify how much currency is entitled to us (or) how much balance is left in the cryptocurrency.

There are many different types of cryptocurrency wallets such as desktop, online, mobile, hardware and paper. The security levels is wholly dependent on the wallet type being used in the transaction. Every wallet contains an address which is used by other users to send cryptocurrencies to that specific wallet. The private key present in the wallet is primarily responsible for creation of digital signatures in order to make the transactions official. For example, when a person is sending you cryptocurrency of any type, they are basically relieving their rights over the coins and then sending it to your wallet's address. In order to unlock the funds, the private key that is present in the wallet must match the public address that is assigned to the currency. It both the private key and public key are equal, the balance of your wallet will increase and the balance in the sender's account will decrease respectively. One of the steps involved in blockchain transactions is mining. Mining is the process of adding new blocks which contain transactions to the blockchain. Miners are the people whose main job is to validate your transaction and add it to the blockchain network. When people submit transactions to the transaction network, those transactions join the transaction pool. When the transactions are submitted, it stays in an unconfirmed state. Miners use these unconfirmed transactions as data that has to be officially added to the block. But there is a barrier to this. In order to gain access over a block, the miner must solve a computational problem called the proof-of-work (POF) algorithm. If the miner is successful is solving this algorithm, the block, containing the transaction data, can be directly added to the network.

4) Publish/Subscribe (PubSub) Network Layer: The PubSub layer is a cloud service that provides publish/subscribe networking functionality to the system. These layers are used to send messages between the nodes in the network, it is the messaging/communication layer of the system. It archives this by using a channel dedicated to specific functions in the network. Nodes are listening to these channels through custom listeners for their respective channels, and every channel has channel-specific instructions on how to deal with any incoming message.
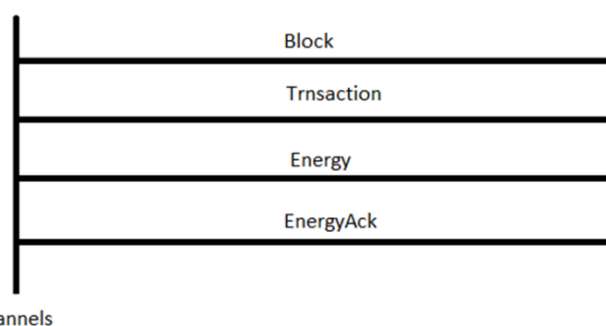
Fig.3 Pub/Sub Network Layer

The system's PubSub layer have the following channels:

1. Block - This channel is designed to send a block and handle an incoming block. The published block function serializes the block into json and broadcasts it. Whereas the block listener deserializes the incoming json block to a block object and adds it to the blockchain, given it passes through all the checks.

2. Transaction - This channel is designed to send a transaction and handle an incoming transaction. The publish function serializes the transaction object into json and broadcasts it to the channel. Whereas the listener deserializes incoming json back into a transaction object. Then it is added to the transaction pool given it passes all checks.

3. Energy - This channel is designed to pass energy transaction requests. The publish function serializes the energy transaction object into json and publishes it. The listener deserializes the energy transaction into an object and adds it to the energy transaction pool.

4. EnergyAck - This channel is designed to send energy transaction request acknowledgement. The publish and listener serialize and deserialize the object, to and from json respectively.

5) Smart Contract: The smart contracts are used in this system to ensure that the peer the get the energy they paid for while the other peer get the coin

promised. Unlike physical objects, energy cannot be transferred immediately from one node to another. There is transmission delay and storage's charging capacity. There are multiple factors that are involved in the transmission of energy. In order to handle there enounces of energy transmission, smart contracts must be used. When one peer agrees to sell energy, it local metering system will prevent the user to access that amount energy and it will be queued for transmission. Until the entire energy is received by the other user, the coin of the buyer will be locked and won't be accessible to either of the peers. This is done to prevent double spending on the buyer side and violation of the contract by hardware substage on the seller side.

6) 3-way handshake in decentralized network: The 3-Way handshake protocol is the phenomenon used by the Transmission Control Protocol (TCP) for establishment of successful connection between two devices before the process of data transmission takes place. Only after the successful connection has been established, transfer of data is possible among the devices. The connection can also be terminated using the 3-way handshake protocol. The connection is done to reserve the CPU and bandwidth of the devices for proper communication. Therefore, it is highly necessary to close the connection once data transfer is completed. The process is as follows:
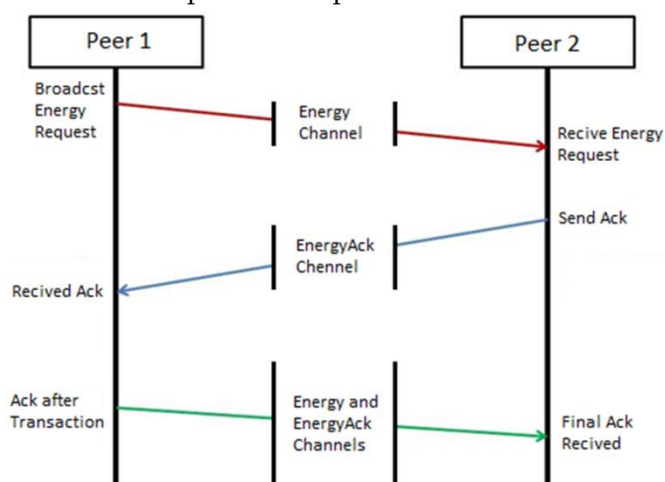
Fig. 4 3-Way Handshake Mechanism

The three-way handshake is used in the network to establish a connection between the buyer and the seller. There are two channels in the PubSub layer dedicated to this specific task; one to broadcast the energy request and one to broadcast in the acknowledgement of the request. Once the transaction is complete, a closed energy transaction is broadcasted in both the energy transaction channels.

IV. DESIGN

A. Design Goals

The design goals of this system are as follows:

1) Primary: The primary goals are as follows:
   a. Decentralised energy trading system.
   b. Better single point of failure resistance in average case than traditional energy distribution software for traditional grid.
   c. Suitable for renewable energy sources like solar and wind.
   d. Remain in control of the users and not a corporation or the government.
   e. Secure network with the latest in cybersecurity techniques.
   f. Every module should go through white-box testing and black-box testing.

2) Secondary: These are the secondary goals.
   a. Create a seamless energy trading platform.
   b. Fluent and modern user interface made using the best and latest frameworks.
   c. Aim to reduce long-term cost for individual users.
   d. Better data pipeline for data collection that allows the data to be analysed for making data driven decisions.

3) Peer Instance Design: The backend of the system contains the code to create and maintain the blockchain and cryptocurrency. Every local node needs to be running python with the environment same as the one used in the project development. Then a copy of the backend will be running on this

python interpreter with the appropriate libraries with version control software for updates. The node will be port forwarded and allows it to exist on the internet.

B. Overall System Architecture
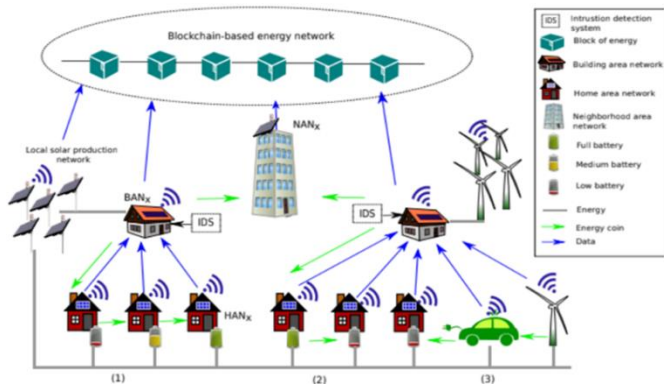
Fig.5 Overall System Architecture

1) Home Area Networks (HANs) does the selling of the surplus production to the electricity grid (other peers seeking enery).
2) Home Area Networks (HANs) has array of solar panels installed on their roofs which are connected smart meter which in turn connects it to the smart grid of energy production.
3) Electric vehicles will act as energy storage devices.

The seller uploads the amount of energy they want to sell, whereas the buyer get look for potential seller to buys the desired amount of energy for a price (paid using EnergyCoins). This is achieved using a publish/subscribe model; the seller publishes amount of energy they want to sell, whereas the buyer subscribe to the "energy channel" and connects with the peer who want to sell energy.

C. Flowchart

1) UI Flowchart:
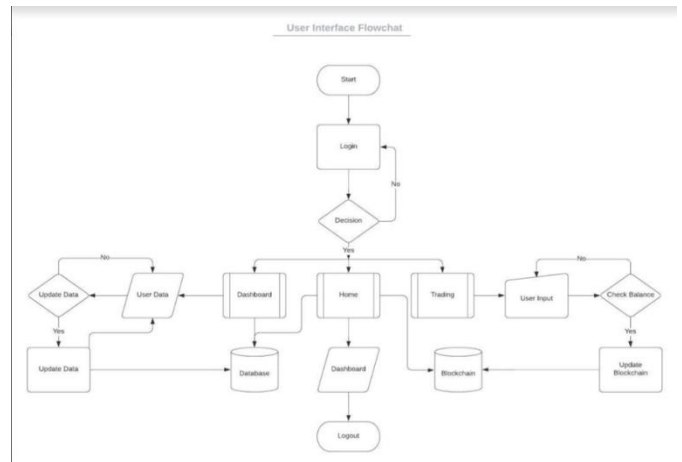
Fig.6 UI Flowchart

2) Server Flowchart:
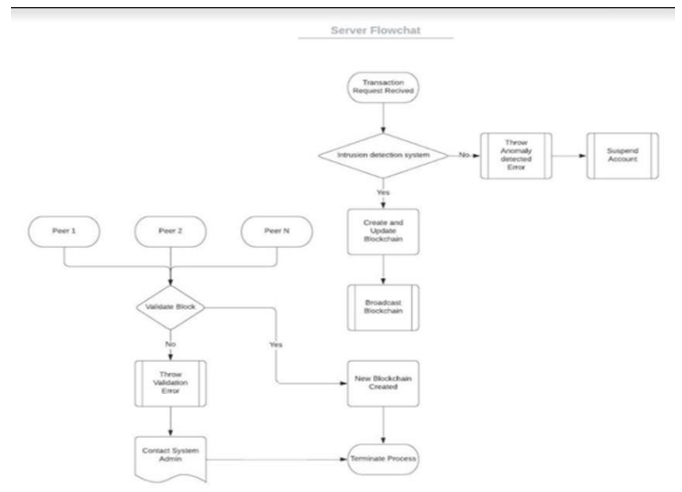
Fig.7 Server Flowchart

3) Check Balance Flowchart:
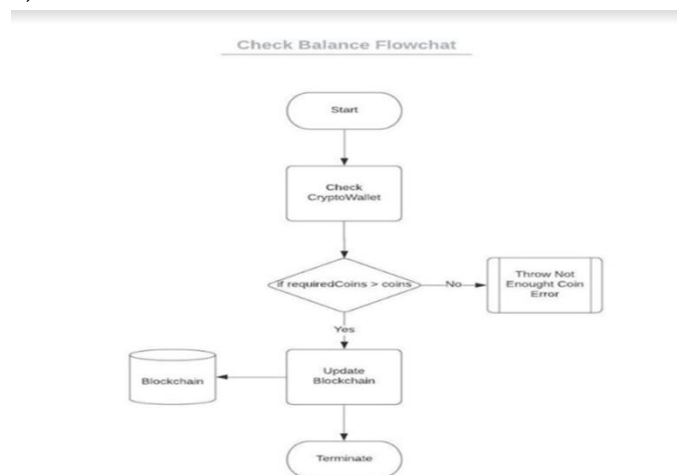
Fig. 8 Check Balance Flowchart

4) Blockchain Architecture Diagram:
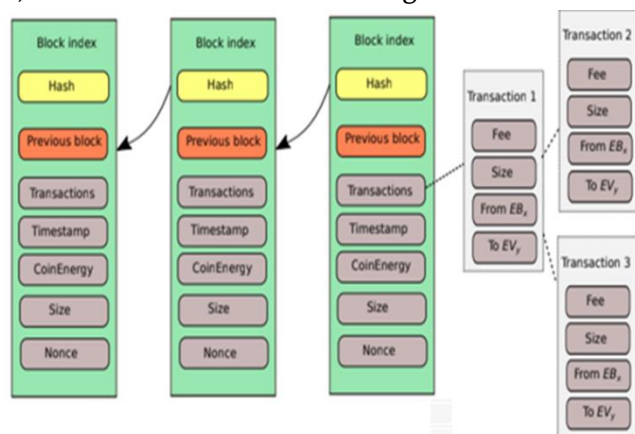
Fig.9 Blockchain Architecture Diagram

V. CONCLUSION

The goal of this project is to create a proof of concept but also a deployable system. This project aims to create a better way of distribution which has the potential of making renewable energy more reliable and in the long run cheaper. This system is very heavily dependent on the underlying custom hardware, economic conditions, political scenarios and public perception. But the system is also necessary to fight climate change as it poses an existential threat to humanity. It will buy us enough time to come up with better energy sources and let the environment heal. Solutions to the problem of climate change have to come from all fields of Science Technology Engineering Mathematics (STEM).

This project put two most cutting edge technology; blockchain and deep learning to help in solving one of humanity's greatest problems. In the process of doing so, it also improves the energy grid and democratizes it. An intelligent smart grid like this one, is out of the control of any single entity.

The project is built using software engineering principles like the agile model. The backend system is unit tested using a module named pytest as it is written in python. Whereas the front-end of the system is written in HTML, CSS, JavaScript and React.js.

VI. ACKNOWLEDGMENT

VII. REFERENCES

[1]. Bitcoin: A Peer-to-Peer Electronic Cash System, 2019, Satoshi Nakamoto Mylrea, M. and Gourisetti, S.N.G., 2017, September. Blockchain for smart grid resilience: Exchanging distributed energy at speed, scale and security. In 2017 Resilience Week (RWS) (pp. 18-23). IEEE.

[2]. B. Roy and H. Cheung, "A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network," 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, 2018, pp. 1-6, doi: 10.1109/ATNAC.2018.8615294.

[3]. C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," in IEEE Access, vol. 5, pp. 21954-21961, 2017, doi: 10.1109/ACCESS.2017.2762418.

[4]. Beigi-Mohammadi, N., Mišić, J., Khazaei, H. and Mišić, V.B., 2014, June. An intrusion detection system for smart grid neighborhood area network. In 2014 IEEE International Conference on Communications (ICC) (pp. 4125-4130). IEEE.

[5]. M. A. Ferrag and L. Maglaras, "DeepCoin: A Novel Deep Learning and Blockchain-Based Energy Exchange Framework for Smart Grids," in IEEE Transactions on Engineering Management, vol. 67, no. 4, pp. 1285-1297, Nov. 2020, doi: 10.1109/TEM.2019.2922936.

HeaBot+ - AI ChatBot for Personal, Emotional and Medical Assistance

Aishwarya Kadali[1], Lakshmi M S[1], Dr. R. Jaya[2]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Assistant Professor Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Healthcare is extremely important to lead a better life and generally patients are not aware about various treatments or symptoms regarding any particular disease. In case of minor problem, the user has to personally visit the doctor for check-up which is usually time consuming. In case of complaints regarding appointments handling telephonic calls are quite hectic. The patient tries to search information about the symptoms or disease and is hard to navigate and arrive at the precise and correct information. To overcome these limitations and problems the solution is a medical chatbot dedicated completely for healthcare. The AI based medical chatbot is the proposed idea which will diagnose disease through a conversation similar to doctor-patient and supply details of the disease and classify the symptoms as minor or major. In case of major symptoms, it suggests the doctor to consult and using machine learning algorithms predicts disease based on the symptoms mentioned by the patient. The bot also gives emotional guidance through conversation using sentiment analysis. The main aim is to reduce healthcare costs and improve accessibility to healthcare knowledge by providing a personalized diagnosis and right health protection.

Keywords: Artificial Intelligence [AI], Medical Chatbot, Symptoms, Disease Prediction, Healthcare

I. INTRODUCTION

Conversational AI is the set of technologies behind the most important automated messaging and speech-enabled applications that offer human-like interactions and conversations between computers and humans. Conversational AI uses various technologies such as Natural Language Processing (NLP), Advanced Dialog management, and Machine Learning (ML) to understand, react and learn from every interaction or conversation. Chatbots are very engaging, asynchronous, cross-channel, data-driven, adaptable, cost effective and prompt. While this technology is still in its developmental phase and

with the progress, health chatbots could potentially increase access to healthcare, improve doctor–patient and clinic–patient communication. The chatbot technology allows for such activities as specific health surveys, setting up personal health-related reminders especially for medicines, pre-diagnosis of disease, booking appointments, medicine recommendation, retrieving and analysing health data.

Therefore, considering the benefits of chatbot, develop AI- based Chatbot which empowers patients by providing them comprehensive list of diets, exercises and other practices. Also supports emotional health through quick conversations. Provides information about the medicines and alternative

medicines and replenishes the stock automatically. Sets reminders for patients to take their pills especially for the elder patients and schedules regular health check-up. Uses machine learning technology to power symptom checker.

II. LITERATURE REVIEW

This paper proposes an idea of medical chatbot with the purpose of building the language gap between user and health providers by replying immediately to the queries submitted by the user. Using techniques such as Natural Language processing the chatbot becomes an entity which imitates human discussion or just replicates a person's health discussion. The aim of the bot is to facilitate to receive a common health related query and either predict the disease with the symptoms or give proper guidance for healthy living or diagnosis. The proposed methods are support vector machine (SVM) algorithm which is powerful classifier which is used for predicting the disease based on the symptoms and Natural Language Processing for three analyses, understand natural language i.e., parse the subject into objects of the sentences and then text description is done and later semantic interpreting which uses knowledge of the word meaning [1]. In this voice-based AI chat bot, a user can give a voice command which will be converted into text. These texts are then sent to an algorithm, which receives the texts as characters and store them as a string to the assigned string. The input of the patient is in Speech format, and then these speeches are converted into text format to verify it with database to give expected output [2]. The next paper elaborates chatbot as software which is able to conduct one-on-one conversations with patients and evaluate each patient's individual requests. The proposed system is a web application with a chatbot within it. The patient enters the symptoms in the chat interface and the system predicts the diseases and also provides daily health tips to the users. The system uses pattern matching algorithms to process the

information therefore the patients get a quicker solution and can act promptly during critical conditions. The data is stored in a knowledge repository and when user issues a query the data is fetched based on the algorithm. This Dialog system uses Natural Language Processing analyses the data based on two variants: rule-based approach and self-learning. The other features include effective symptom-based disease prediction provide remainders about the appointment and as well suggest doctors when needed [3]. This paper explains about Natural Language Processing. Automation manipulation of natural language like speech and text by a software system is Natural Language Processing (NLP). In every communication either verbally or in written whatever we express carries huge amount of information. Some type of information can be interpreted and value can be extracted from the topic we choose to communicate, our tone, our selection of words in communication. NLP gives the machines the ability to read, understand and derive meaning from this information [4]. This signifies the important enhancement in Chatbot and various techniques available for designing chatbot for healthcare. Chatbot are described as stateful services and recalling preceding information so as to give functionality. The proposed system uses Artificial Intelligent Markup Language to train model and Microsoft voice synthesizer used for identification of words spoken by the patient. This bot gives composition of the medicines prescribed and their dosage needed. The methodology used is word order similarity between sentences which uses tokenization and lemmatization and NLP which involves text scheduling and text comprehension [5]. This AI Chabot uses dialogue design system to help users to reach the correct diagnosis, logic for state transitions are made, natural language generation templates were used, and system initiative to the user and get responses from the user. This chatbot has three main conversational phases: acquisition of basic information, symptom extraction, and diagnosis. User enters a loop of symptom

extraction states until it acquires sufficient information for a diagnosis. Users have the choice of entering the loop again to talk to the doctor about another set of symptoms after be given their first diagnosis and the other choice is that the user can read their history of chats about what they have discussed [6]. The next paper has validated the proposed interactive paradigm to a real clinical context where the chatbot has the goal of recommending several disease prevention pathways. The proposed system has given importance to data generated from numerous flows and showing Big Data features from several resources. The problems such as data sensitivity of medical records, massive data handling and context scalability led to this study and design and realise systems and architectures based on big data analytics and combing this technology with the Chatbot interface to provide efficiency and scalability [7]. The process of drawing meaning from text is called Semantic analysis. This technique allows computers to understand and interpret sentences, paragraphs, or whole documents, by analysing their grammatical structure, and identifying relationships between individual words in a particular context [8].

III. EXISTING SYSTEM

The existing system is that the patients have to physically consult the doctor and wait for the appointments which are tedious in case of early symptoms and due to this many tend to neglect their health henceforth leading to complications. Health organizations connect to their patients using health applications for behaviour coaching or medication monitoring which lack engaging user experience and affects scalability factor of technology. The other existing system of engaging patients is through telephone helpline systems as a part of customer service which is tedious job in terms of financial sense and customer satisfaction.

In case of Web-based customer service application patients are required to fill enquiry form with personal details and drawback with this is that patients are not willing to disclose their personal information to customer services option if it doesn't provide security features. Another risk with a Web-based system that doesn't rely on live chat is infuriating consumers with delayed responses. And if all the above means does not work for the patient/customer, he may have to pay a visit the hospital or the medical institute to get the information he is looking for.

The problem with the personal relationship between a patient and their physician is that it is generally conducted in an 8–10-minute consultation, which is not a very satisfactory. Moreover, doing Google searches to find the right answers is also challenging as most patients do not know how to assess the quality of information they find online. In traditional psychotherapy, users will meet a counsellor at a specified time and place usually at the therapist's office and speak face to face. Users can have a real-time conversation, in which the practitioner can pick up on the body language and tone of voice to get the insights on what the user is feeling.

IV. PROPOSED SYSTEM

The proposed system is an artificial intelligence based chatbot which can answer health query posted by the user, and conducts a conversation similar to a patient doctor conversation. With the information received from the user using Natural language processing techniques the relevant response is given. Using the support vector machine algorithm, the disease is predicted based on the symptoms entered by the patient during the conversation. The necessary health tips and diagnosis procedures are made available to the user and if the disease is classified as major the necessary doctors are recommended and even appointments are booked. The chatbot can even send remainders in case of medicine monitoring and the other details of the medicines like the composition or dosage are also provided by the chatbot. The chatbot

also provides emotional assistance through the conversation using semantic analysis and asks the user to lead a healthy and happy life. The chatbot is built-in within a web application.

A. User Logins to Web Application

User is required to register and login to the application to access the medical chatbot.

B. User Asks Medical Queries

User can ask general medical queries related to diseases or the symptoms related to a disease.

C. Queries about Medicines

User can ask queries related to medicines. When the medicine name is given the details about the medicine is given to user.

D. Disease Prediction based on Symptoms

Based on the symptoms given by the user the gradient boost machine learning model predicts the disease.

V. DATA PREPARATION

A. Sources and Observation

For creating the data for disease prediction, we can take the help WHO and kaggle. WHO provides diseases dataset which has the diseases names and its symptoms. For medicines we collected the data from WebMD and Drugs.com. These dataset have all details of drugs, like its side effects, effectiveness, condition in which it should be used, age limitation etc.. For understanding mental health of users, we collected twitter data to gain the knowledge about user's emotion and to differentiate between each expressions and emotions. For diet planner, to suggest a diet we need to have a dataset which has calories, protein, fat, carbohydrates information of each ingredient.

B. Quality and Filtering

The data is explored and checked for its consistency for proper working of the chatbot. Diseases have many symptoms and signs and for each disease the severity of symptom can vary and many diseases have some common symptoms which need to be taken care.

To avoid the biasness in data, the quantity of data for each disease should be in equal. All symptoms have certain idioms which have to be linked to common language for better understanding of symptoms. The chatbot reverts with a set of questions to answer when the symptoms given by users are not sufficient for proper diagnosis.

For critical or chronic conditions, the chatbot recognizes and categorizes these conditions into major disease and suggests users to consult doctor immediately, it recommends the doctors who are available and close to user's location when user gives their pin code. Upon confirmation from user to consult doctor the chatbot schedules an appointment. There are some symptoms which are recognized and help the chatbot to classify the conditions as major disease. The symptoms which have higher severity rate are spotting urination, burning micturition, patches in throat, swelling of stomach, swelled lymph nodes, chest pain, weakness in limbs, pain in anal region, irritation in anus, stomach bleeding.

C. Pre-processing

The data is pre-processed to make it consistent for the classification. The pre-processing techniques which are used are text pre-processing and Label Encoder. Text pre-processing techniques include tokenization, lower casing, stop word removal, stemming, lemmatization, normalization, and noise removal. In tokenization the text splits into smaller tokens with the help of delimiter, non-alphanumeric characters such as white space, special character, or punctuations. The stop words like articles, conjunctions, prepositions are removed after tokenization as these words have no information and removing those makes the data more consistent. In lower casing all uppercase characters are converted into lowercase. Lemmatization pre-processing is used to replace the word/keyword with basic word forms. Finally, the data is normalized, unlike lemmatization, normalization typically depends on the task. For example, in normal text, words like "b4" and ":)" were

normalized to "before" and "smile". But in this task the data is about medicines, drugs, disease, dictionary mappings is done for these data.

VI. PROPOSED METHOD

A. Natural Language Processing

Natural language processing (NLP) is the heart of AI driven chatbots and integrating NLP means adding a more human touch. NLP forms the basis of language recognition and empowers technology to understand natural language speech and text-based commands. It includes mainly two components: NLU- natural language understanding and NLG-natural language generation. Using sophisticated NLP algorithms chatbots can process the queries which include interpreting, inferring and determining what was meant and then define a series of appropriate actions. NLG involves mapping the given input and analysing various aspects of language, text planning and text realization whereas NLU aims at converting and handling unstructured data into structured form understandable by the system. For the chatbot to understand the message NLP follows the five main steps i.e. lexical analysis, Syntactic analysis, Semantic analysis, Discourse integration and pragmatic analysis. NLP faces the primary challenge of understanding the complexities of human language.

B. Gradient Boosting Classifier

Boosting is a Machine Learning algorithm in which the weak learners are converted into strong learners. Irrespective of distribution of training data weak learners are classifiers which always perform slightly better. The predictions are sequential in boosting, each subsequent predictor learns from the errors of the previous predictors. Gradient Boosting is a technique of generating additive predictive model by combining various weak predictors, i.e., Decision Trees. Figure [1] represents how actually gradient boosting classifier works.

As initial prediction we start with a leaf which represents initial prediction of every individual. For classification, this will be equal to log(odds) of the dependent variable. Residual for each observation is calculated to build a decision tree which predicts residuals of each leaf. With the predicted residuals, calculate new probability to obtain new residuals. This process is continued until the residual value is close to 0 or the number of iterations matches the value given as hyper parameter while running the algorithm. After the calculation of output values for all trees, the final log(odds) predicts the disease based on given symptoms. The final computation is calculated by the following formula.

Initial prediction + Learning Rate × Predicted Residual₁ + Learning Rate × Predicted Residual₂₊… …

Learning rate scales the contribution of each tree, and there will be a trade-off between learning rate and number of trees. Usually, the used values of learning rate lies between 0.1 and 0.3.
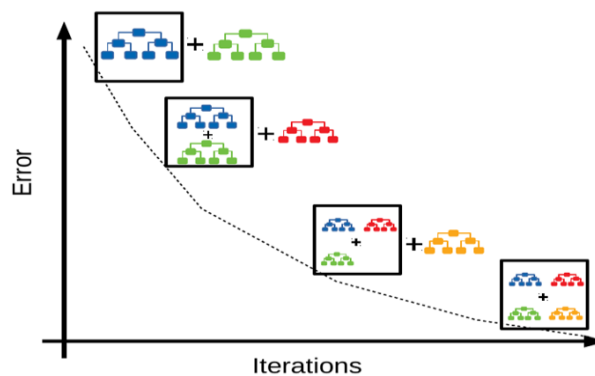
Fig. 1 Gradient Boosting Classifier

C. Support Vector Machine (SVM)

A supervised ML model which uses classification algorithm for multi-group classification problems. In this project SVM is used for emotional text classification. This algorithm is used along with text classification problem. This model is given sets of labelled training data for each category and later the model is able to categorize new text.

The algorithm works by taking the data points and outputs a hyperplane that best separates the classes. The line is called the decision boundary.

For SVM with natural language processing we are using simple SVM classifier and TF-IDF to create the model to classify the text.

VII. EXPERIMENTAL RESULTS

This work performed on datasets of different sizes and different symptom severity rate for each disease to demonstrate how well the developed model can work on data of all sizes and different severity rate of symptom. We are using two datasets of different sizes with different severity rate of symptom to predict the disease. We trained the model with 80% of dataset whereas remaining 20% is taken for testing. Training and testing time differs for every algorithm, so the result of accuracy.

Decision tree and gradient boosting classifiers showed same accuracy rate for smaller dataset and the accuracy rate was higher when compared to logistic regression, random forest and naive bayes classifier. For larger dataset, logistic regression, random forest, multinomial naive bayes classifier, decision tree and gradient boosting trees showed same accuracy to a point. But when validated with different kfolds, Gradient boosted tree was the most efficient algorithm with K=2.
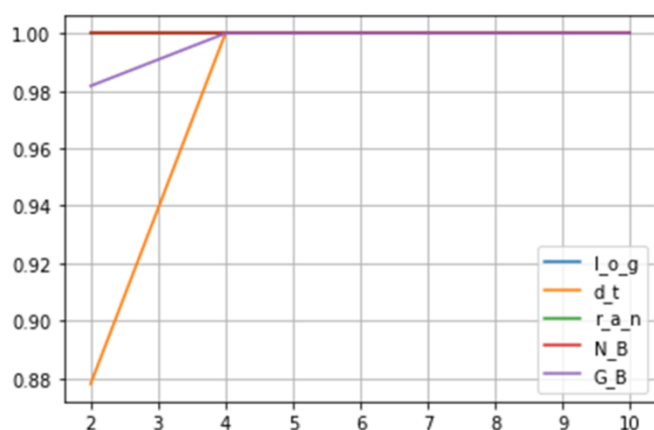
Fig. 3 KFold Validation of algorithms

VIII. ADVANTAGES OF CHATBOT

1. Medical chatbot makes self-care easier for the users by acting as a virtual assistant and providing timely medical advice.
2. Chatbot provides the consumers an interactive platform that is engaging and personal, quick and easy to access, works in real time, effective, responsive and inclusive.
3. Through chatbot the user can have informal communication between user and computer providing new opportunities to improve consumer's engagement process and efficiency by reducing the typical cost of customer service.
4. With the help of AI and machine learning algorithms, chatbot are forecasted to save healthcare costs when used in place of a human or assist them as a preliminary step of helping to assess a condition and providing self-care recommendations.
5. The conversational AI chatbot helps the patients to input their queries and facilitate instant responses. Patients can start the conversation with the chatbot and it will be saved in the database for future reference. The chatbot will classify the user symptoms with series of queries and symptom conformation. The chatbot classifies the disease as major or minor disease to emphasize the seriousness of disease so that appropriate medical diagnosis or treatment is taken.

IX. CONCLUSIONS

The main aim of the project is to bridge the gap between users and medical information using the medical chatbot. The future is the era of internet. In healthcare, Internet enabled many people to access to information that can aid in the diagnosis of health conditions or the development of suitable treatment plans. To progress to that, this AI based healthcare chatbot enables user to communicate with conversational agents through voice and text.

The chatbot response to the users' queries using the natural language processing techniques. The

knowledge base stores information relating to the medical keywords, remainders, frequent system responses, logs and customer feedback information. As each module is very complex and has variety of functionalities therefore each module can be optimized and each module must be developed in such a way that it is upgradable independently. The chatbot also provides support for multiple disease prediction using different machine learning algorithms and also classifying the disease severity. The user is also allowed to post queries about medical dosage and information on medicines which can limit the browsing time and produce accurate information to the user. Thus, this medical chatbot will give medical assistance to the users in case of mild symptoms or when doctor is unavailable hence improving the efficiency and performance of the bot through various conversations.

X. FUTURE SCOPE

The future era is the era of messaging app where people spend more time in messaging and require only relevant information quickly therefore medical chatbot has huge and large future scope. For the project medical chatbot, by taking the advantage of extensibility of the system can develop the chatbot with voice and face to mimic a counsellor and also interact with the patient at deeper levels. This can extend for the people with disabilities when the user gives voice input the system can generate a voice-based output. As the system is scalable each module can be upgraded by adding more functionalities and increasing the medical information content in the knowledge base. The efficiency of the chatbot can be improved by adding more combination of words increasing the medical terms due to which the medical chatbot could handle all types of diseases. The future scope is to build a vast repository of reliable healthcare information to offer relevant answers to all patient queries. The chatbot can be made available in multiple languages as this is

limitation for this project which impacts the user experience.

The other important future aspect is content. A lot more contextualization is required to carter the needs of a very large audience. For this many data points need to be captured and chatbot needs to be rigorously trained. In future, we can develop and provide AI- powered chatbot to monitor the health conditions of user for all kind of diseases. Users can customize the chatbot to monitor their health for specific disease condition like cancer, Thyroid, Maternity etc.

In the future, the chatbot's symptom recognition and diagnosis performance could be much improved by adding support for more medical features, for instance location, duration, and intensity of symptoms, and more detailed symptom description. A research from mental health charity Mind shows, at any time, one in out six workers are experiencing common mental health problems which includes anxiety, depression and other mental health issues. So, in future this chatbot can be integrated into a wider web application for screening and providing support to employees who are working with disturbing online media in the workplace and to users who are using social media platform.

XI. REFERENCES

[1]. Navida Belgaumwala, Dr. Rajashekarappa, "Chatbot: A Virtual Medical Assistant", International Journal for Research in Applied Science & Engineering Technology, June 2019.

[2]. Flora Amato, Stefano Marrone, Vincenzo Moscato, Gabriele Piantadosi, Antonio Picariello, and Carlo Sansone," Chatbots meet eHealth: automatizing healthcare", Workshop on Artificial Intelligence with Application in Health, Italy, November 2018.

[3]. Mrs. Rashmi Dharwadka, Dr.Mrs. Neeta A. Deshpande, "A Medical Chatbot", International

Journal of Computer Trends and Technology (IJCTT), June 2018

[4]. Rashmi Dharwadkar, Dr. Neeta A. Deshpande, "Pharmabot - A Recommendation on General Medicines– Survey", International Journal of Innovative Research in Computer and Communication Engineering, June 2018.

[5]. Muse Mohamud Mohamed, Professor Wang Zhuopeng,"Artificial Intelligence Healthcare Chatbot System", International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE), February2020 .

[6]. Ashwini Shangrapawar, Ankita Ravekar, Sakshi Kale, Nidhi Kumari, Aman Shende, Pankaj Taklikar, "Artificial Intelligence based Healthcare Chatbot System", International Research Journal of Engineering and Technology (IRJET), February 2020.

[7]. Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill,Cherie Armour, Michael McTear, "Towards a chatbot for digital counselling", Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI), July 2017.

[8]. Chin-Yuan Huang, Ming-Chin Yang, Chin-Yu Huang, Yu-Jui Chen, Meng-Lin Wu, Kai-Wen Chen, "A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion", International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, Thailand, December 2018.

Whip-Smart

Deepthi S[1], Deeksha S[1], Harini M[1]

[1]Department of Computer Science, new horizon college of engineering, Bangalore, Karnataka, India

ABSTRACT

Whip-Smart is an E-learning application designed for the kids of 2-7 years of age. Usually the kids of these ages do not have a good attention span. To grab their attention and to be able to educate them can be a difficult task for both parents and teachers. This age forms the most important part of their life as the foundation for future studies is based on what they learn in their early years. Hence it becomes more important to educate the kids in the best possible way and make them understand every concept very clearly. This application is designed keeping in mind the same views about children's education. The children of different ages are assigned with a series of day-to-day tasks pertaining to their age that they're supposed to complete. On successful completion they're rewarded and their performance throughout is recorded. The detailed report of the child's performance is available for the parent to view. The application is designed in such a way that it is user friendly. The tasks are designed with animations and interactive learning is ensured to make it more interesting for the child. Various areas of study for pre-school learning are covered. Data integrity and security are ensured. Due to the pandemic situation and lockdown in various parts of the world, the most affected are the kids of these ages as they require more care and attention while learning. And in various parts of the world where availability of good education is in itself a huge privilege, a pre-school education tool like this serves the purpose of providing quality education and preparing the children of today for a better future with good knowledge and understanding.

Keywords: Kids, E-learning, CNN, Machine Learning, Data Analysis

I. INTRODUCTION

Intelligence is not rare among human beings. It is found in children at birth. With the proper stimulation, it's possible to nurture the event of reasoning and problem-solving skills in young children. The first six years of life are the most important years of a child's education. One of the greatest struggles people face as parents is how to foster digital savvy- ness in their children – while they don't want to keep them away from today's latest technology, they would want to make sure that

this is the application that they do not have to feel any guilt in handing it to their child when the going gets tough, because it undoubtedly stimulates educational (and sometimes even emotional) growth, and not to mention that it is extremely fun and super engaging while also ensuring that it is safe and secure. What they experience will instil curiosity in children and help their little brains grow.

Parents can always use some extra help at times to make their children learn and sometimes that help is an app. With the advancements in technology, finding information on how to enhance and track

your child's development is literally at your fingertips. This interactive application offers various development activities which will help you track your child's development. These categories include motor skills and emotional, sensory and cognitive development; all focused on helping your kids reach his/her first milestone. All activities in this app are engaging and also there are beautiful rhymes and songs.

Children's brain development happens rapidly from the time of birth up to three years. There are four main areas of development: motor (physical), language and communication, social and emotional, and cognitive development. Brain development is part of cognitive development. This cognitive development tells us about how a child's intellect grows, and that includes thinking, learning, understanding, analysing and problem-solving skills. These skills play a prominent role in all the other areas of development of a child. Hence the first three years of a child's life are very vital for learning and overall development of the child.

Due to the pandemic, all the schools are shut and the students are stranded at home and education is at a standstill. The older children who are learning via online classes have the ability to understand the concepts even without much guidance, but what about the kids? The Application Whip- Smart will always help kids in their overall development. This app aims to assist you in simplifying daily life and be the best parent or guide through each stage of your little one's development. Other features can also be included such as a personalized daily calendar of your baby's development, information on your child's health and safety and weekly activities for your child using the data obtained from the interaction with this app.

II. SYSTEM ANALYSIS

A. Existing System

There are various online learning platforms for kids like ABCmouse, Reading Eggs, Starfall, Brain pop for kids of the age 2-8 of years, which concentrate mainly on teaching a lot of science, math and other knowledge through videos which is great but does not focus on overall development of the child. Moreover, the assessments to analyse the performance may not be accurate. With the help of data analytics we will be able to analyse the child's performance and interests and also provide a report to the parent.

B. Proposed System

Our platform Whip-Smart makes use of the latest technology to provide the child with the best learning experience. It is majorly focused on providing certain features to enhance the child's visual perceptions and assimilation of knowledge. The main focus is to make learning fun and interactive for kids. Some of the various features we are trying to implement are handwriting recognition (to teach the child to write digits/characters), cartoonifying images (to teach the real time objects) , expressive text to speech recognition used for storytelling. The curriculum is carefully designed in the best interest of the child and its progress is carefully tracked and a detailed report of the child's progress and interests are available for the parents to view. Several machine learning algorithms and data analytics framework are used to achieve the above features.

The platform aims to teach the children various topics like numbers, addition and subtraction, shapes, patterns, measurements, and more. The games and artistic activities make math fun while giving young learners the practice they need to create a robust foundation for fulfilment. The unique drawing and painting program can provide children of all ages the chance to use lines, shapes and colours to make original works of art. Games, books, paint-by-number activities, and puzzles can teach children to memorize and use a variety of colours. The curriculum in our platform includes several highly entertaining and interactive activities, each teaching a specific learning

topic. These animated videos engage children while they learn. There are many jigsaw and cut-out puzzles designed to help develop problem solving and critical thinking skills. Puzzles also help to aid the children in remembering important ideas and skills in reading, math, science, art, social studies and music.

III. HARDWARE AND SOFTWARE REQUIREMENTS

A. Hardware Requirements

Processor   :   500 MHz Processor
RAM   :   512Mb
Hard Disk   :   10 GB
System type   :   64 – bit OS

B. Software Requirements

Operating system : Windows XP or Higher
Front End   : React.js
IDE   : Visual code
Framework   : Django
Data Base   : SQLite

IV. DIAGRAMS
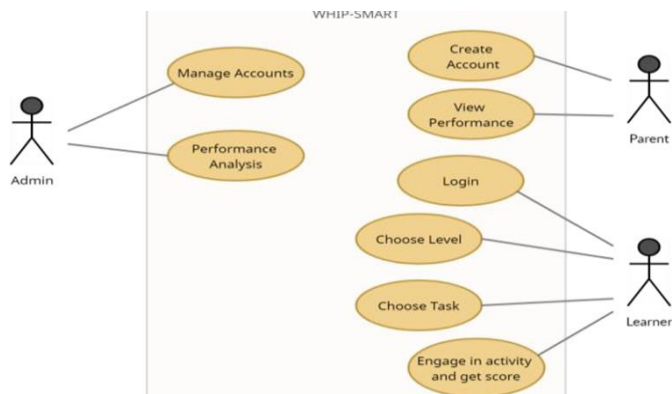
Fig.1. Use Case
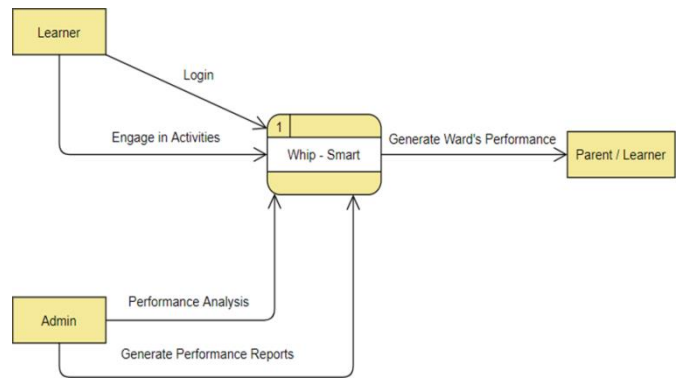
Fig.2. Data Flow Diagram (Level 0)
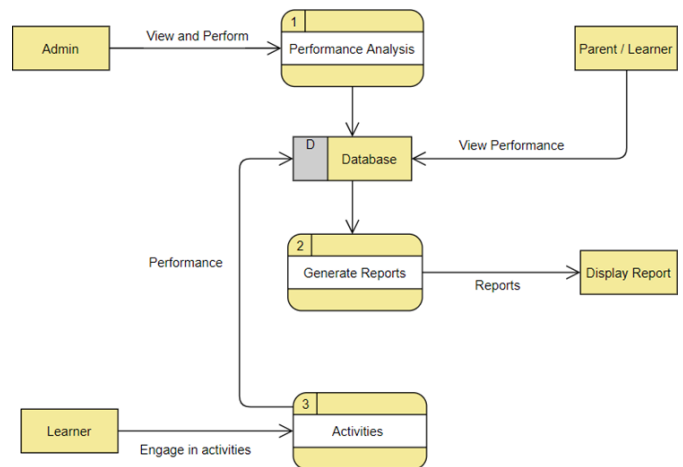
Fig.3. Data Flow Diagram (Level 1)

V. SYSTEM IMPLEMENTATION

A. UI Implementation

The UI is designed using reactjs. There are different levels designed to develop cognitive, listening, reading and writing skills. Parents can view their ward's performance and actively participate in their learning journey. React sketch library enables children to practice the digits. It captures the image format of the written digit and sends it for digit classification implemented using CNN.
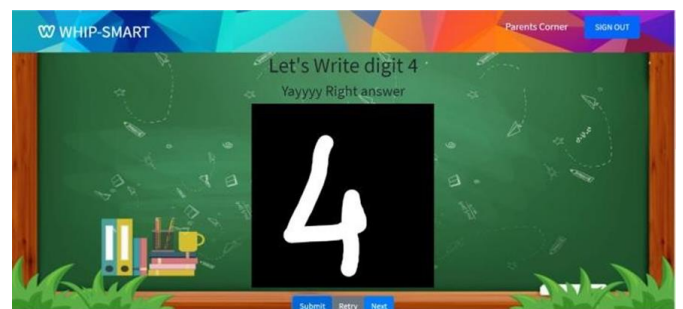
Numbers, alphabets, animals, stories, rhymes and so on are read aloud for the children using the text to speech synthesis which can be implemented using the library available in react called 'react-speech-kit'

B. Algorithm Implementation

The image that is saved in the frontend using the react- sketch library of Reactjs can be fed to the backend (Django) using Axios where the digit is classified using the CNN model. Convolutional neural networks work great for images/data that can be represented in grid format. We cannot directly feed the image to CNN model; it needs to be converted into the right format before we feed it to the neural network for classifying the digits.

After pre-processing the image as mentioned before, we need to train the model using the model.fit() which is available in the keras library of python. Once the image is successfully classified it is sent back to the front end where the result is displayed.
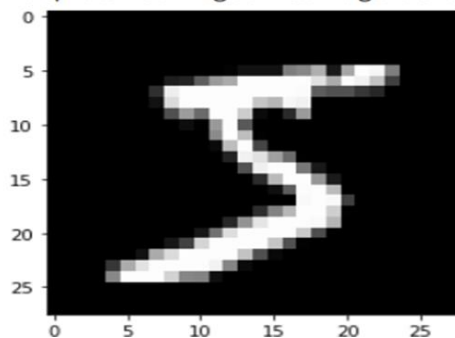
C. Accuracy

Accuracy of the built model can be found by using accuary_score which is available in sklearn.metrics.y_test and predicted variables which store the prediction are passed as a parameter to accuary_score.
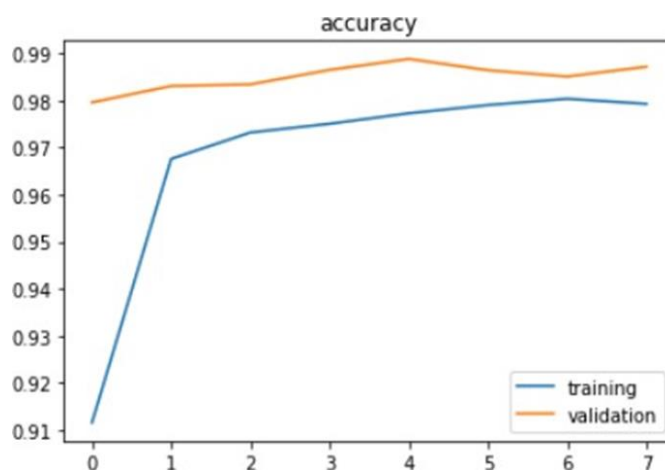
VI. CONCLUSION

This project/software proposes to build a user friendly platform for kids to pursue their early education. It is always quite a difficult task to get the attention of kids to teach them. But by using our application the learning process for kids is made easy. There are various functionalities implemented to cater to the kid's busy mind and make learning fun and interactive.

VII.   FUTURE SCOPE

This application mainly aims to provide quality education to children between ages 2-7. It proposes to build a user- friendly platform for kids to pursue their early education. This application can be further extended for ages beyond 7. Based on the kid's ability to comprehend the concepts, the level of complexity of questions can be auto generated. Every child has a unique level of understanding and it is important for an e-learning platform to identify this and cater to the child's needs accordingly. The current education system has a pre-defined curriculum which is usually not updated for a very long time and is generalised on the whole.

The child's progress can be tracked and recorded and the same report is available for parents to view. The child's interests and strengths can identified via Classification algorithms in Machine Learning. This

app aims to assist parents to simplify their busy lifestyle without compromising on their little one's development. It can also feature a customized daily calendar of the child's development, information on the child's health and customised weekly activities for the kid.

VIII.   ACKNOWLEDGMENT

This project and the research behind it would not have been possible without the guidance and support of our guide Dr. Rachana P. We would like to thank her for monitoring the project development and for constantly motivating us. We would also like to thank all teachers from New Horizon College of Engineering, Bangalore.

IX. REFERENCES

[1]. Olisah Kingsley S, Mohamed Ismail Z "Web Based E-learning System for Pre-school Kids .2020". [online] Available at: https://www.researchgate.net/publication/319360 301_Web_Based_Elearning_System_for_Pre-school_Kids.

[2]. Jyoti Shinde, Chaitali Rajput,Prof. Mrunal Shinde, Prof. Milind Rane "Handwritten Digit Recognition" vol 2 issue 2. IJTRSD publications. Ijtsrd.com.2020.   [online]Available   at http://www.ijtsrd.com/papers/ijtsrd8384.pdf

[3]. Glad Chinda "How To Build a Text-to-Speech App with Web Speech API" Published onDecember 12, 2019

[4]. Convolutional Neural Network [online] Available at: https://deepai.org/machine learning-glossary-and-terms/convolutional- neural-network

A Smart Water Regulating System Using Internet of Things

Balakrishna Gudla[1], S S P M Sharma B[2]

[1]Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, Karnataka, India
[2]Department of Mechatronics Engineering, Parul University, Vadodara, Gujarat, India

ABSTRACT

At present circumstances IoT and distant recognition technological procedures have deployed in varied streams of research for direction, controlling and evaluating data from remote locations. Slurp water is the necessary component for everyone, but its efficacy is tested on regular basis during the rendition process. This intricacy arises with the explanation related to limited water resources, expansion of occupants, and quick development of essential offices. A wide necessity in the development in exercises to framework the administer water attributes. Some extent there is disparity in water where people's and animals' health will suffer as a result of these characteristics, which will also have an impact on the physical environment's stability. In order to ensure the protection of slurp water the it is necessary to consider a standard which is maintained properly. Water contamination is a possibility the best fear for the green globalization. Potentials should be built in an innovative manner in order to safeguard the protected and secure reserves of drinking water. In the suggested system, we present a smart and efficient system that is simple to configure for water regulation assistance in the Internet of Things. The module is equipped with a set of transducers that are used for calibrate the water's variables. Temperature, pH, turbidity, and stream transducers of the aqua can all be calculated as variables for a design [11]. The controller can create the standards from the transducers. As a controller, the Arduino model can be used. With the help of a WI-FI module, the transducer data can be viewed on web servers at any time.

Keywords — Internet of Things, Temperature, pH, turbidity, stream transducers

I. INTRODUCTION

In the present scenario there were a lot of innovative concepts. evolving, At the same time, however, there is a undesirable effects, There are gradual temperature differences over the world and more, so there is dependable enough water for the earth's impacts. In the current situation water peculiarities monitoring in present circumstances Featuring challenges as a result of environmental temperature variations, water assets, occupant expansion, and a variety of other considerations. As a result, there is a pressing need for progress in order to effectively monitor the water nature criteria in current conditions. [1].

Provisions for water The attention of hydrogen ions is determined by pH. It demonstrates whether the water is acidic or alkaline. The pH scale ranges from 0 to 14. It should have a pH range of 6.5 to 8.5 for drinking purposes. Turbidity is a measurement of the amount of large aggregates of long particles in water that are not transparent.. If there is a lot of turbidity, there is a lot of danger, such as diarrhoea and cholera. [11]. If

there is decreased turbidity, we can conclude that the water has been purified. Temperature transducers are used to calibrate the level of water, whether it is cold or hot. In the suggested system, water flow transducers were used to compute the direction of flow[11]. Aqua monitoring systems have traditionally been linked to physical accumulation of aqua snippets from remote locations [2].

The Internet of Things is a noble archetype that integrates transmit with various devices or utilities using a transducer and the identification of appropriate Arduino Microcontrollers [11]. The suggested system uses the Internet of Things to implement a smart water control system (IoT). The proposed system is used to keep track of the pH, temperature, and turbidity of the water. With the ability to create web applications connected with microcontrollers, the Internet of Things plays a critical role in extending the resolves to numerous executions. [3].

II. RELATED WORK

Current Situation is considered as an ideal stage of tainting, steady augmentations in temperatures, unstable and imperiled wellness confusions. Water pollution is the most extreme problem existing known to man as of now, which is basically harming streams or coastline. Water corrupting occurs at a condition where noxious substances are catapulted right away or accidently into water framework. Water polluting will impact seeds and beasts alive inside the water framework. Furthermore individual strength is vainglorious by pollutants of water framework. Water spoiling is a foremost entanglement which involves existent assessment and adaption of water resources overseeing hypothesize at the phases of worldwide to autonomous also. It has been analyzed that water pollutions are most significant states of downfall and contaminations all around the world. In our nation obviously 580

residents lapses each day because of water pollution issues [1].

Around the worldwide the statistics says that almost 14000 people are influenced by water debasements every day. In different quickly developing nations corrupted water is being sent for savoring the nonattendance of certified on time investigation. A remarkable intention in this marvel is obliviousness of individuals and authority and insufficiency of water qualities checking system produces critical wellness confusions. Common conditions will likewise change the norm of water. As water is most fundamental component of living life forms it is amazingly compelling to make sure about it [2].

The proposed model plans a transducer system which helps to administer the conduct of water by the data distinguished by the transducer lowered in the water framework. Employing numerous transducers, this method is able to assemble abundant factors from acqua, for instance, temperature, pH and turbidity [3]. The quick headway of distant transducer organizes and development gives a novel method to manage steady data acquiring, transmission and taking care of. The clients can get advancing water attributes information from distant. By looking above issues, we made and organized a negligible exertion water quality checking system that can screen water quality ceaselessly using IOT condition. In our proposed system water quality boundaries are assessed by the particular water quality noticing sensors, for instance, pH, turbidity, and temperature. These sensors are dealt with microcontroller. The readied data can be seen through a program application using a one of a kind IP address. In addition, with the help of IOT condition, we can offer office to get to data distantly from wherever all through the world [4].

III. DEVELOPMENT OF SYSTEM

In the cutting edge time as there is simplicity to get to innovation with extensive reach with the bountiful accessible assets IoT has taken of a transcendent

feature with utilization of logical business, technique [6]. This advancement is typified with huge scope in intelligent things, structures and transducers misuse types of progress in preparing power, devices downsizing, and arrange interdependencies make an offer brand-new capacities not previously possible. Internet of Things modules like coordinated automotive systems, shrewd traffic systems, and transducers embedded on roads with frameworks attract us closer to "splendid metropolitan networks", which help limit stop up and essentialness usage.
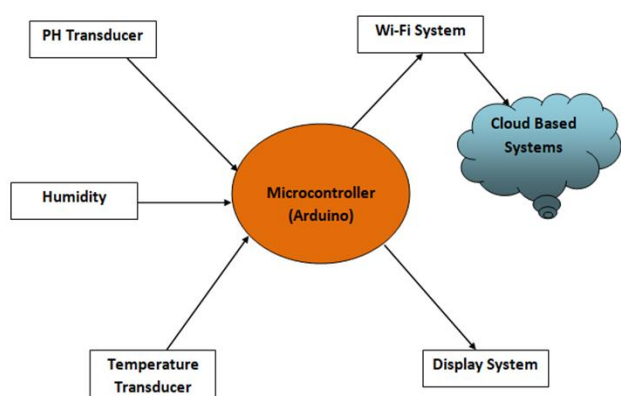
Figure-1: Representation of Proposed System

This proposed module, we present the theory on consistent checking of water attributes in IoT condition. The complete cycle of proposed framework is explained. Consistently module of cycle is expounded to sum things up. The proposed system includes various transducers pH, Temperature, turbidity, stream, Humidity) are combined with Atmega 328 regulator. The microcontroller is utilized to deal with the transducer information and taking care of them to pass on data by means of web. Arduino is utilized as a center regulator with transducer data to be shown with the help of Wi-Fi modules [5].

Arduino can recognize the environmental factors by amassing assets from a various transducers (pH, temperature, turbidity) and commits the data to cloud approved organizations. The microcontrollers are reconfigured utilizing Arduino essential programming

language. The cloud can save essentialness by controlling the contraptions and actuators are fundamentally used to kill on and turn the devices. Online system is used to give the customer receptiveness to work from any territory by considering these components IoT has been progressed which render circuit of devices by deploying internet and employing IP address as character. At the point when it is connected to Wi-Fi Module of the microcontroller makes unmistakable IP address [6].

A. Temperature Transducer

To measure the demeanor of DS18B20 transducer is utilized. DS18B20 transducer aligns force of warmth by and large than an adequately utilized thermistor since it is business personality transducer. It induce taking off yield possibilities than thermocouple so no convincing motivation to strengthen the yield potential. The yield potential is decisively with respect to the Celsius temperature. The DS18B20 transducer is available in a pre-wired and waterproof form. While the transducer has a range of 125℃, the connection is jacketed in PVC, thus it is recommended that it be kept below 100℃.[11] The degree of hotness or frigidity of water framework is determined by the power of warmth in water. [7].
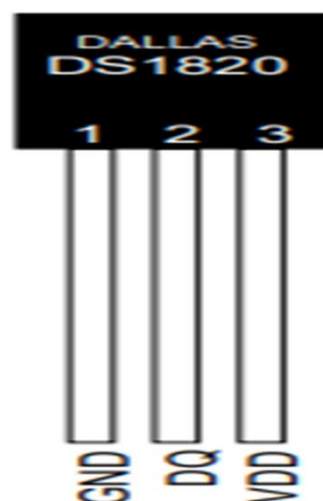
Figure 2: DS18B20 Transducer

Attributes:

- Power Supply range is 3v to 5.5v
- No stand by power is required
- Temperature range is -55°C to 125°C
- Accuracy of ± 0.5°C
- No extrinsic modules are required

B. pH Transducer

The hydrogen molecule's obsession is measured by pH, which refers to the force of hydrogen. The degree of corrosiveness or alkalinity of a panacea is measured by its pH. The pH scale is a logarithmic scale that ranges from 0 to 14, with a neutral point of 7.[11]. Values above 7 indicate a simple or basic panacea, while qualities below 7 indicate an acidic objective. It takes away at 5V for a long time. [8].
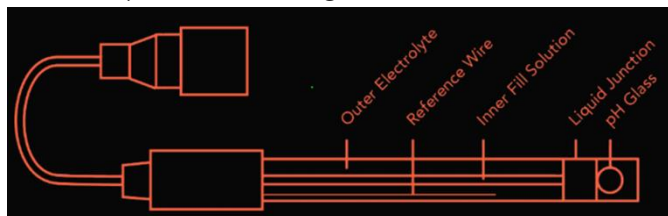
Figure 3: pH Transducer

When the anode potential is decreased in the panacea, the modifying hydrogen particle immersion has no effect on the anode potential. Through a hub, a panacea at the referred to terminal connects with the illustrating panacea and the processing cathode, attaining the organisation. The yield of the aligned cathode varies with temperature (despite residual activity at constant pH), As a result, a temperature transducer is required to precisely account for this change in yield. This is accomplished by the use of an analyzer or a transmitter. [9].

C. Turbidity Transducers

Turbidity is a measurement of water's opalescence. [11]. Turbidity is a term used to describe the force of warmth that causes water to lose its clarity. It has been determined to be a sufficient calculation of equipped water.[11] The required brightness of dive water flora is limited by turbidity. It is also utilised to

raise the water surface temperature over the standard because the enthralment of the level of hotness from sun sparkle helps to raise the water surface temperature over the standard. [10].

Figure 4: DS18B20 Transducer

Turbidity is an internationally recognized criterion for assessing the qualities of drinking water, and a turbidity estimating device is a combination of logic and hardware with light that assesses turbidity by calculating the scattering of brightening emitted rapidly through a water example, obliging colloidal specks that asylum in fection. The Nephelometric Turbidity Unit (NTU), also known as the relative Formazin Nephelometric Unit, is the most reliable way to determine turbidity (FNU).[11]. Nephelometry is the process of aiming a light discharge at a liquid sample and measuring the force of light scattered at 90 degrees to the shaft. [5].

D. Ethernet Shield W5100 (Web Server)

The Ethernet Shield is a low-cost, high-sensitivity Ethernet shield that may be used at 5 volts. Ethernet shield is open source, and writing PC programmes with the Arduino Uno as a guide is difficult. W5100 provides an IP-based system that supports both TCP and UDP, as well as PHP, HTML, and other programming languages, and can be used as a tool for a variety of Application Program Interfaces. We can quickly mean regional workers who can be secured from a division by readdressing the switch's standard IP by transmitting W5100 Ethernet shield. [11].
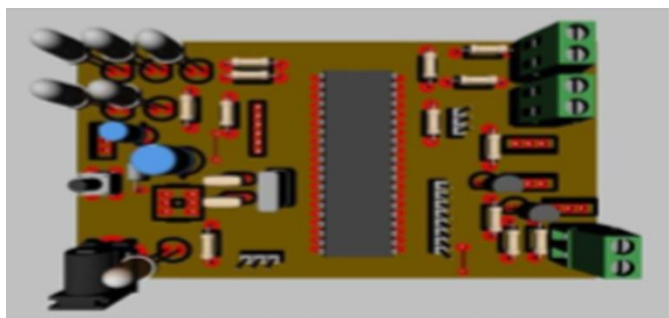
Figure 5: PCB Layout

E. Wi-Fi Module

The ESP8266 Wi-Fi Module is a free SOC with a constructed TCP/IP show stack that can provide access to your Wi-Fi coordinate to any microcontroller.[11]. The ESP8266 is programmed to either support an application or offload all Wi-Fi frameworks organisation constraints from another CPU. An AT course set firmware is pre-installed on each ESP8266 module. The ESP8266 module is a staggeringly monetarily wise board with a monster, and reliably creating, network [3].

Figure 5: Wi-Fi Module (ESP8266)[11]

F. Arduino Uno

The Atmega board is connected to the PC through a USB port, which allows for simple interface. A system with Windows-based movement, Macintosh-based movement, or even Arduino-based movement can be connected to the Arduino-based board. [11]. A free mode indicates that the power connector is tied to the force on the contraptions, and that the PC will, for the most part, draw power from itself. It has a 14-stick plan with six basic pins (0-5) and 14 input/yield pins (0-13), all of which may protect simple

information sources.[11]. It has 5v and 3.3v power connectors, as well as a ground connector. [4].
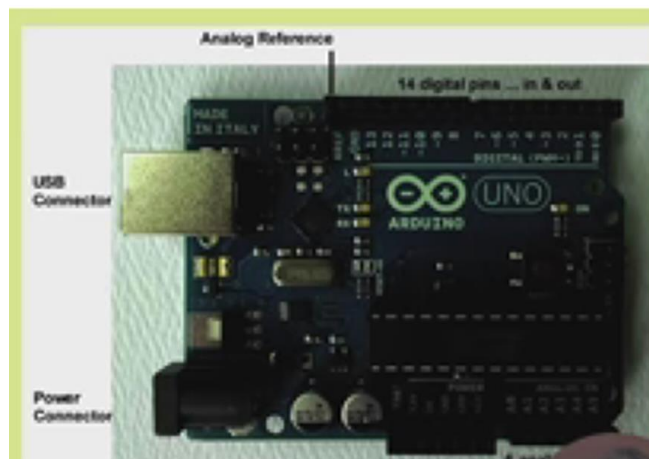
Figure 7: Arduino Board [11]

IV. FLOW TRANSDUCER

A transducer is used to monitor the flow of water in a stream. A plastic valve body, a rotor, and a Hall Effect transducer make up this transducer.[11]. When water/liquid passes through the valve, the pinwheel rotor rotates, and its speed is proportional to the flow rate. Every time the pinwheel rotor is rotated, the Hall Effect transducer will produce an electrical heartbeat. [2].

V. PROPOSED IMPLEMENTATION

The proposed model may have the ability to execute the task in a way that does not take into account present structures in various elements such as computerised mobility and easy-to-use affiliations.[1]. In every practical sense, the suggested system is made up of the control circuit, Wi-Fi Module, Microcontroller, and web workers (PhP). The Arduino Microcontroller and pH/temperature/stream/turbidity transducer are used to control the model. [11].

The machines in the proposed system can similarly examine the water while following and observing the framework, which will aid in abstracting the unsafe of swallowing water. The proposed framework is

capable of realising various states of contraptions that have the openness of divergent subsystems that are linked to temperature, stream, and pH will be deemed IoT used utensils. [2].

The proposed system is based mostly on IoT, which has recently sparked thought in global development.[11]. There are generally two areas covered: the first is equipment, and the second is programming language modification.[11]. The transducers in the equipment division will aid to measure continuous functions, while the Arduino remakes basic attributes to discrete ones, and the PC displays the yield from the transducer, while the Wi-Fi module provides the interaction between equipment and programming.[11]. In programming, we created a subject to introduce the C programming language. [4].
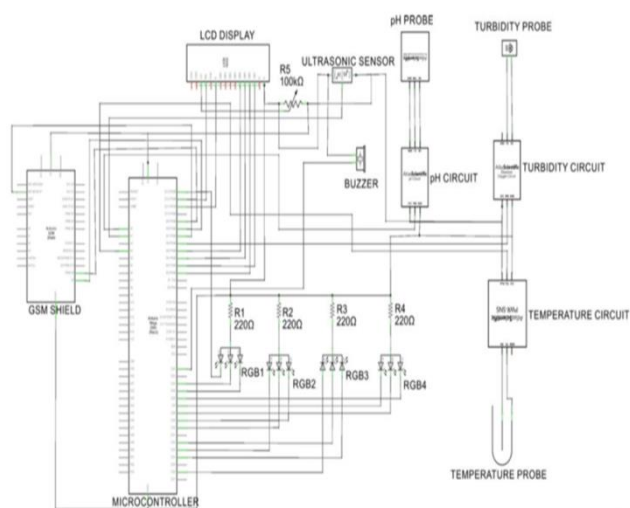
Figure 8: Circuit Diagram

Transducers for turbidity, stream, pH, and temperature are included in the equipment pack. The thickness of water is measured using a turbidity sensor. Suspended solids (mainly soil particles) and small fish (minor plants and animals) suspended in the water section provide thickness.[11]. Low levels of turbidity may produce an audible sound. [11]. The Water Flow transducer can be used to measure liquid movement in mechanical and nearby applications. PH stands for Power of Hydrogen, which refers to the size of the hydrogen molecule. The temperature of the water is regulated and displayed using a

temperature transducer. The Arduino regulator receives the yields of the transducers. It is utilised to convert continuous data into discrete data, and LCD visualises the outputs from transducers.[11]. The Wi-Fi module provides the connection between the equipment and programming language. [7].

C Language is used to explain the product module of this suggested framework. The PCB is depicted during the critical phase of erection, with fasteners such as transducers affixed to it. To observe the result, webservers are created and accessed in the android version.[11]. When the organisation begins, dc current will be supplied to the module, and Arduino and Wi-Fi hubs will be turned on. [11]. To demonstrate framework, the water factors are examined and their outcomes are mentioned. Using a similar methodology, after the module is installed on a piece of determined water framework and Wi-Fi is enabled. [11]. The proposed framework can explore its present time limits on a PC/cell phone from any location at any time. [9].

VI. RESULTS

A. Experimental Setup

Water checking and oversight framework for swallow water usage comprise of far off recognizing gadgets that are set in underwater domains to gather data, for instance, stream, temperature, pH and turbidity.[11]. The assembled information are imparted to distributed computing gadgets by methods for Wi-Fi empowered correspondence model [4].

The entire module arrangement was created because it consists of microcontrollers connected to transducers and electrical devices that allow all machines to be controlled. [5].

The developed module can be tested using the instructions for identifying units and launching Ethernet/Wi-Fi-based frameworks in identified locations. [11]. The stated framework is consistently used and gives consistent representation by means of motorization information by connecting

Ethernet/Wi-Fi to the Internet Protocol with the aid of IoT apps. [8].

As a result, the water testing exercises have been presented as IoT for enhanced remote usage and site oversight. The proposed model identified a capable execution model with four transducers and many modules, all of whose functionalities have been cleansed. The ATMEGA 328 with Wi-Fi module is used in this model in the proposed framework. The introduced contraption is connected to the internet via an ADC and Wi-Fi module built in. Transducers are connected to the Arduino UNO board for testing; the ADC will convert the corresponding transducer examining to its mechanical worth, and the contrasted characteristic boundary will be calculated from that value.
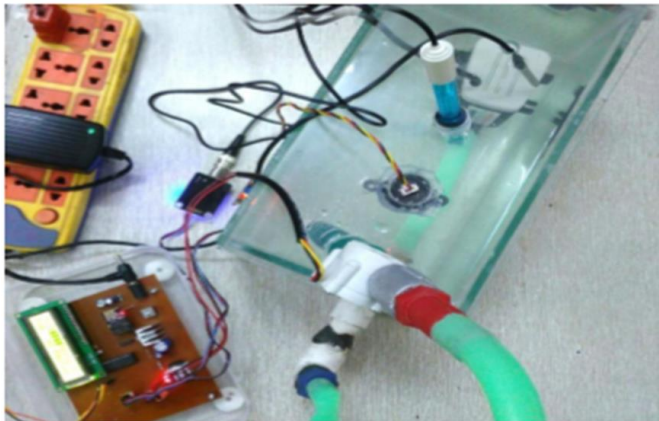
Figure 9: System Setup

During the time spent separating data from various transducer devices, which are set expressly region of interest. When a proper affiliation with a cut off device is established, the recognised data will be sent off the web worker. [9].

Figure 10: The turbidity-voltage relationship

B. Server-based monitoring devices

Typically, a microcontroller is interfaced with Ethernet to create a specific IP, and a page is created for which IP is created in the URL, after which the page will execute and the customer can manage any electric devices by turning them on and off. [11].

Figure 11: Page for logging in
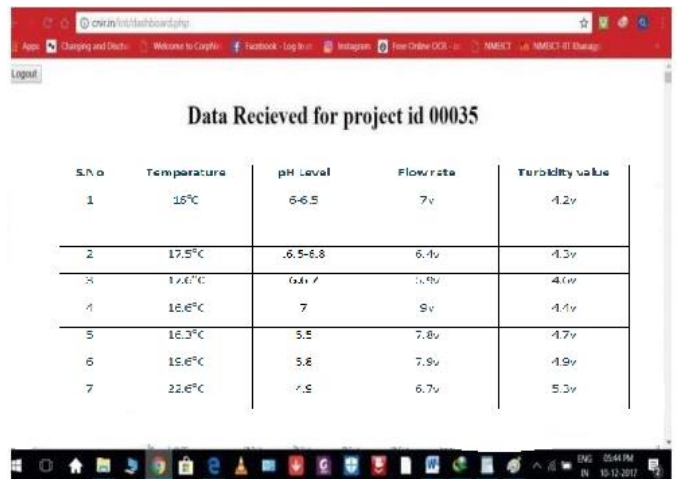
Figure 12: Personalization of Profiles

VII.   CONCLUSION AND FUTURE SCOPE

With the current improvements, it is expected that web availability would be widespread and at a phenomenal level. Transducers and gadgets are successfully connected, and collaborative data can be accessed from anywhere in the globe [5]. With the help of exercise, non-basic disappointment adjustment, and plausible energy, the internet of things (IoT) has simplified the alignment ability to web work boosted adroit utensils solidly and persuaded arrive at data capacity everywhere.[6]. The proposed framework reduces the specific availability of energy. The hard and fast eagerness eaten by

gadgets can be reduced by attaining cloud based examining and observing of framework through IoT by stimulating the proposed programmable organisation on an easy to use module. [7].

Managing the turbidity, PH, and temperature of water conveys using a transducer with a specific level heading and running Wi-Fi framework. The module can manage water properties without the need for human intervention, and it is less expensive and does not require human intervention. So, aside from being modest, appropriate, and quick, the water trademark is deserving of attention. The company is quite adaptable. This module can be used to regulate various water boundaries by simply substituting similar transducers and changing the required computer routines. The procedure is straightforward. The firm can be contacted for hydrologic, noxious air effect, manufacturing, and farming production, among other things. It has a broad appeal and a large amount of money. [8].

By possessing the installed devices in the circumstances for controlling approves mind assurance (i.e., shrewd climate) to the climate [2]. To utensil this prerequisite to utilize the transducer contraptions in the conditions for social event the data and studying. By utilizing transducer devices in the environmental factors, the proposed framework can bring the circumstances into existing situation for example it can speak with substitute elements through the framework. At that point the assembled data and investigated results will be open to the customer with client accreditations through the Wi-Fi [4].

The achievability of the proposed network is to produce an ostensible endeavor in commitment and movable partnership contraptions for blending alluringly Internet of things with far off analyzing framework [5]. An utensil utilized at particular territory of the globe can have the choice to manage from other region of the globe. Various gadgets may have the ability to communicate with one another over an unspecified time period using

indistinguishable insurgency [6]. By 2030, it is expected that IoT-based executions would encourage the growth of increased work advancement and city regions, and as previously stated, it will have security insufficiency, which should be addressed urgently. [7].

VIII. REFERENCES

[1]. N. Kedia, "Water quality monitoring for rural areas- a Sensor Cloud based economical project," 2015 1st International Conference on Next Generation Computing Technologies (NGCT), 2015, pp. 50-54, doi: 10.1109/NGCT.2015.7375081.

[2]. Jayti Bhatt and Jignesh Patoliya, "IoT Based Water Quality Monitoring System", International Journal of Industrial Electronics and Electrical Engineering, pp. 44-48, April, 2016.

[3]. Michal lom, ondrej priby and miroslav svitek, "Internet 4.0 as a part of smart cities", Smart Cities Symposium Prague 2016, pp. 1-6, 2016.

[4]. Zhanwei Sun, Chi Harold Liu, Chatschik Bisdikia, Joel W. Branch and Bo Yang, "Water Quality Monitoring System In The South Sudan", 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2012.

[5]. S. Kartakis, W. Yu, R. Akhavan and J. A. McCann, "Adaptive Edge Analytics for Distributed Networked Control of Water Systems", 2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), 2016, pp. 72-82, doi: 10.1109/IoTDI.2015.34.

[6]. Mithaila Barabde and shruti Danve, "Real Time Water Quality Monitoring System", International Journal of Innovative Research in Computerand Communication Engineering, vol 3, pp. 5064-5068, June 2015.

[7]. Akanksha Purohit and Ulhaskumar Gokhale, "Real Time Water Quality Measurement System based on GSM", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE),

Volume 9, Issue 3, pp. 63-67, Ver. V (May - Jun. 2014).

[8]. Eoin O'Connell, Michael Healy, Sinead O'Keeffe, Thomas Newe, and Elfed Lewis, "A Mote Interface for Fiber Optic Spectral Sensing With Real-Time Monitoring of the Marine Environment", IEEE sensors journal, vol. 13, no. 7, pp. 2619-2625, July 2013. DOI: 10.1109/JSEN.2013.2258760

Rainstorm Prediction System

Ms. Harshitha H[1], Ms. Pooja Kumari[1], Ms. Simran Agarwal[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Road, Panattur post, Kadubeesanahalli, Bengaluru, Karnataka, India

ABSTRACT

Rainstorm is a devastating disaster that usually occurs during rainy seasons at Himalayan regions. The recent floods in the _Kedarnath' area, Uttarakhand are a classic example of flash floods in the Mandakini River due to cloudburst that devastated the country by killing thousands of people besides livestock. The traditional methods used for cloudburst prediction are weather forecasting, data mining techniques for weather prediction by modelling meteorological data, laser beam atmospheric extinction measurements from manned and unmanned aerospace vehicles. These techniques are more expensive and time consuming along with uncertainty of accurate prediction. The proposed method in this paper is Arduino based cloudburst predetermination system with real time calculation of rainfall intensity.

The rainfall prediction is done with the use of machine learning in minimal costs. The complete weather forecasting setup is flexible enough to be installed anywhere and make weather predictions without much historical experience. We used different machine learning algorithm to check the accuracy of rainfall prediction.

Keywords— Aerospace, Disaster, Forecasting; Intensity, Prediction, Arduino.

I. INTRODUCTION

Cloudbursts are extreme weather events in which an area registers more than 100 mm rain in just one hour. So, create a prediction system that will be able to forecast extreme weather events such as cloud bursts at least two days in advance.

Early prediction and warning of such severe local weather systems is crucial to mitigate societal impact arising from the accompanying flash floods.

Farmers can know when to plant or harvest their crops.

People can choose where and when to take their holidays to take advantages of good weather Surfers known when large waves are expected Regions can be evacuated if hurricanes or floods are expected Aircraft and shipping rely heavily on accurate weather forecasting

This architecture is compared with other previous proposals and it demonstrates an improvement on the ability to predict the accumulated daily precipitation for the next day.

II. MOTIVATION BEHIND THE RESEARCH

Today's world the enormous data sets required and inherent unpredictability of the Earth's atmosphere makes predicting future events very tricky indeed. Current computer models are required to make judgments of several large-scale phenomena. These

include things like how the Sun heats the Earth's atmosphere; how pressure differences are affect wind patterns and how water-changing phases (ice to water to vapor) affect energy flow through the atmosphere.

One example is the Numerical Weather Prediction (NWP). This model studies and analyses vast data sets from satellites and other sensors to provide short term weather forecasts and long term climate predictions.

Complexity of time series data for various cases of weather patterns with more number of parameters can result in quantitative estimation

This system also helps many people to save their lives before knowing the weather in advance.

III. LITERATURE SURVEY

Literature survey is the most important step in the software development process (4). Machine learning technology adapts and responds to data, learning over time to better answer search queries. This helps provide better accuracy in search results at a speed beyond human capabilities.

Cloudburst Predetermination System Arpit Tiwari 1 , S K Verma2 (Department of CSE, GBPEC, Pauri, India)

IOSR Journal of Computer Engineering (IOSR-JCE) Cloudburst is a devastating disaster that usually occurs during rainy seasons at Himalayan regions. The recent floods in the ‗Kedarnath' area, Uttarakhand are a classic example of flash floods in the Mandakini River due to cloudburst that devastated the country by killing thousands of people besides livestock. These techniques are more expensive and time consuming along with uncertainty of accurate prediction.

- **A neural network-based local rainfall forecast system using meteorological data on the Internet:**

A case study using data from the Japan Meteorological Agency

In this study, we evolve and test a native rainfall (precipitation) prediction system based on artificial neural networks (ANNs). Our arrangement can involuntarily obtain meteorological statistics used for rainfall forecast from the Web. Meteorological data from apparatus fixed at a local point is also shared among end users in our system.

- **A neural network-based local rainfall forecast system using meteorological data on the Internet:**

In this study, we evolve and test a native rainfall (precipitation) prediction system based on artificial neural networks (ANNs). Our arrangement can involuntarily obtain meteorological statistics used for rainfall forecast from the Web.

Meteorological data from apparatus fixed at a local point is also shared among end users in our system. The ultimate goal of the work was the practical usage of ―big data‖ on the Internet. Also, the splitting of data amongst end users for precise rainfall prediction.

IV. COMPARATIVE ANALYSIS

In this paper the comparative analysis mainly done based on the temperature, Atmosphere, humidity. Based on the above criteria the below section shows the some of the aspects that were less/not found in the previously existed system.

The existing system produces certain disadvantages;

- It is extremely difficult to accurately predict the rainstormin two or three days(5).
- The traditional methods used for cloudburst prediction are weather forecasting, data mining techniques for weather prediction by modelling meteorological data, laser beam atmospheric extinction measurements from manned and unmanned aerospace vehicles.(5).
- These techniques are more expensive and time consuming along with uncertainty of accurate prediction.
- The back-propagation-feed advance neural network can be utilized in numerous applications such as character identification, weather and financial forecast, face detection etc(5).

- The present paper presents a mechanism of prediction of cloudburst by detecting the formation of cumulonimbus type cloud, using the brightness temperature (TB) (5).
- The proposed system removes some of the drawbacks of existing system such as;

The proposed system provides the prediction in 2 or 3 days advance.

- Unlike traditional system,which predicts the rainstormin 5 to 6 days. Documents with quality rich textual content will be included in the top results.
- This system also helps many people to save their lives before knowing the weather in advance.
- Most of the documents that are produced in the results are most relevant to the user's request.
- It consumes very less amount of time to be implemented unlike other techniques that consume a lot of time to process very huge database and further finding patterns of hidden knowledge in order to produce predictions.

V. PROPOSED SYSTEM

This system will predict rainstorm based on parameters such as temperature, humidity and wind. This system is a web application with effective graphical user interface .System will take this parameter and will predict cloud burst by comparing the present and previous data in database. System will calculate cloud burst based on these data, therefore this prediction will prove reliable.

This also uses the ARDUNO and RAIN GAUGE to get the precipitation of the rain and then it will get the output to the machine leaning algorithms.

Here we are implementing on the basis of machine learning with its algorithms which are suitable for forecasting the rainfall in particular. Depending upon the various factors given above.

We are implementing through APIs of dataset and there we are taking the data for which state we are predicting the cloudburst.

It consumes very less amount of time to be implemented unlike other techniques that consume a lot of time to process very huge database and further finding patterns of hidden knowledge in order to produce predictions.

- Descriptions of data to be entered into the system
- Descriptions of operations performed by each screen
- Descriptions of work-flows performed by the system
- Descriptions of systemreports or other outputs
- Who can enter the data into the system
- How the systemmeets applicable regulatory requirements

A Support Vector Machine (SVM)

This a computer algorithm that learns by example to find the best function of classifier hyperplane to separate the two classes in the Input space.The SVM analyzed two kinds of data, i.e. linearly and non-linearly separable data .The example of linearly separated data is shown in fig. below. Best hyperplane between two classes can be found by measuring the hyperplanemargin and find out the maximum points

$f(x) = w\ x + b(1)\ T$

Where x refers to a training pattern, w is referred to as the weight vector and b as the bias term

Artificial Neural Networks:

Artificial neural networks (ANN) have existed in various forms since the 1940s (McCulloch and Pitts, 1943; Good fellow et al., 2016), but have received renewed interest in recent years (Good fellow et al., 2016). An ANN is a collection of neurons, which are small computational units that superficially mimic the way neurons work in nature. A single neuron is simply a weighted sum of a set of inputs, plus a bias, with an applied activation function, Fig. 1 (left). A non-linear activation function fact(·) is important for

success in applying ANNs, otherwise the resulting model output is simply a linear combination of the inputs. The equation for a single neuron can be written as: yk = fact(b+xiwi) (1) The power of ANNs comes from connecting many neurons together in a network. The simplest network structure is a feed forward network, as shown in Figure 1 (right). Neurons are connected in simple layered structures where the inputs of each neuron are connected to all the outputs of the previous layer. If we describe the inputs xi and weights wi in matrix form, we can write a whole layer of neurons as: y = fact(W x) (2) where the bias is included as w0 = b by adding an artificial constant input x0 = 1,. A feed forward ANN is built by connecting multiple layers together. The inputs to the network are connected to the inputs of the first hidden layer. The first hidden layer can then be connected to more hidden layers. The last hidden layer connects to the output layer. The output of the ANN is given by this output layer. We can then write a single non-linear matrix equation for the whole network. An example equation for an ANN with three hidden layers is:

y out = f out act (Wout f (2) act (W(2) f (1) act (W(1) x

))) (3)

]

Equation (3) shows that an ANN is simply a non-linear matrix equation with a large number of coefficients. Each W(j) matrix can be large, thus allowing the ANN model to fit complex non-linear systems.

Random forest:

Random forest is a kind of machine-supervised learning algorithm on understanding the ensemble. Ensemble training is indeed a form of erudition whereby you multiply combine various kinds of algorithms or just a similar procedure to create a more efficient model of forecasting. This algorithm combines many algorithms of a similar kind i.e.

multiple decision trees, resulting in forest of trees, hence the term "Random Forest". This algorithm may be used for tasks of regression as well as classification. The initial steps involved in applying the random forest algorithm are selecting M arbitrary archives from the dataset, constructing a decision tree largely dependent on those M archives, choosing the amount of trees we want to see in the algorithm and repeating steps one and two. For classification issues, the group to which the new record belongs is predicted by each forest tree. At last, the new record was allocated to the division where the majority vote is received. The algorithm of random forests is also not bias, as there have been various trees and then each tree is proficient on a data subset. The random forest algorithm relies on the influence of "the audience;" hence, the algorithm's common bias reduces.

Polynomial regression

Like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x). Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function E(y | x) is linear in the unknown parameters that are estimated from the data.

Logistic regression

This is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for

various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

The advantages of proposed systemare;

- Rapid and abrupt climate changes have remarkably increased the importance of a weather app.
- Today, weather apps can provide accurate forecasts to enable users to make informed decisions.
- Integration of advanced features and technological progress have made it easy to know the sudden changes in the weather conditions along with real - time alerts.
- Since ages, humans want to know about weather conditions in advance. In the current age of the Internet and mobility, weather apps make it possible to predict weather conditions on the move. Summing up, your weather mobile app with necessary features can make the users prepared for any weather-related

VI. ARCHITECTURE OF THE SYSTEM
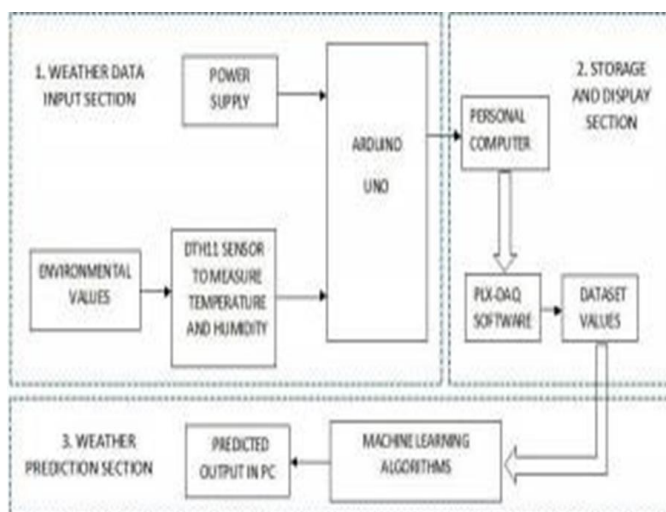
Fig: 1 Architecture of the System

The above architecture diagram shows us the estimated overall architecture of the whole system

A. Weather Data Input section

In this section, a combined sensor to measure both temperature and humidity is used. The sensor works well to send the real time data continuously to the interfaced microcontroller. The controller is powered with a voltage supply after rectification done in a DC adapter. The function of this section is to collect the readings of the weather data. The measured values are ready for further processing based on the requirement.

B. Storage and Display Section

In this section, the output is then given to the Personal computer and then displayed in an Excel file. This is done by means of specific software to interface the Arduino outputmeasured values to excel file. The final sheet of excel will be the dataset file available for the machine learning prediction.

C. Weather/Rainstorm Prediction Section

This section includes the setup for the dataset being given as input to machine learning models to accurately predict the rainfall based on various algorithms. The algorithms will make the prediction based on various approaches and different sets of data. The predicted percentage value is displayed on the Personal computer and can be compared with previous historical weather input

VII. IMPLEMENTATION

The flowchart diagram below clearly describes the working of the system.
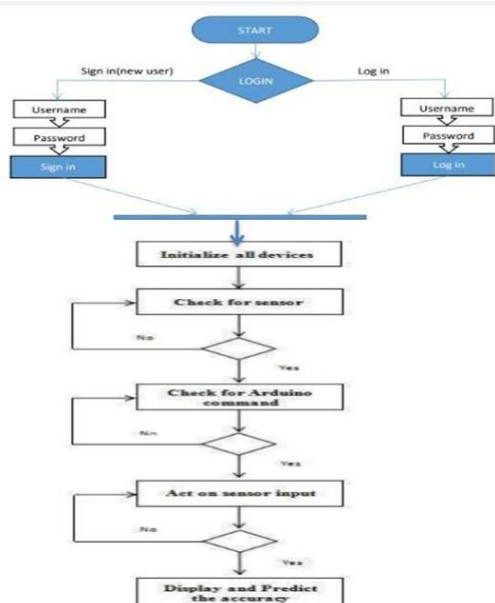
Fig: 2 Flowcharts

So, to have a better understanding of the dataset and for better comparison, first preprocess should be done. Before going for the prediction, preprocess can be done. It is representation of the dataset in form of graph. It eases the process of comparison and along with that it also gives a better understanding of the dataset present. Dataset should be split in two parts, the first part deals with training the algorithm used and the rest part used to predict the amount of rainfall. Rainfall is predicted only with the algorithm with more accuracy. The algorithm used should undergo training before it does prediction. So, in this part of the system, the training is been done. This is done with the above mentioned approaches. This step gives a proper idea of which algorithm is more accurate among the all. Then the remaining dataset (which is not used in training) is being used and rainfall prediction is been done. This part is also done in both the approaches. Finally, after the all the process is completed, the result is received in form of graph and table which shows the future rainfall and the accuracy of the algorithm. The accuracy is received in the form of Metrics and excel sheet. In Metrics along with the accuracy different types of errors are also shown and the same is represented in

the excel sheet. After all, at last the predicted value is stored in excel sheet and is received.

A. Errors calculated

The accuracy of the approaches is being calculated against the types of errors that can produce negative effect on the algorithm. These errors can affect the algorithm's accuracy and hence are being calculated. The types of errors that is being calculated are MAE, MSE, RMSE and RSQUARED. MAE calculates all the absolute errors and then finds the mean value for all. It first calculated the mean of all the dataset present, then subtracts the mean value with each data individually and add all the resultant value and finally divides it with the total number of dataset present. MAE = (1) Next error is MSE. It is almost similar to mean absolute error. MSE = (2) The only difference is, instead of adding the resultant (subtracted value of mean with each dataset), it finds the square of it and add them. RMSE error is being calculated by subtracting all the predicted and actual values with each other, finding all the squares of it and adding all the squared value the total value that we will receive is stored. The stored value then further divided by total values present. The resultant value is squared rooted. (4) (X= regression error (sum squared) Y = total error (sum squared)) the above errors are being calculated by subtracting the division value of sum squared regression error and sum squared total error with value one.

B. Front end

This screen will pop up in the screen as soon as the project starts. The user needs to give the login for using the application. If he/she doesn't have they can register with the new email id and password and can login so they can use the application successfully.

C. Anaconda

Anaconda is an open source software used to run python codes.

Jupiter notebook, spyder

The errors values and the accuracy value are popped up in anaconda with all algorithms. In anaconda, while the project is being executed, the dataset which is being used is shown there. So, while execution of the code, the user can cross verify it and can stop the processing instant on finding that the wrong dataset is being used.

D. Visualization

This option deals with representing the dataset in form of graphs. Different types of graphs are being produced after execution of this process. Below are the different Metrics and graphs:

Fig 1 explains the Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them.

Fig 3

Fig 4

The accuracy score achieved using random forest to compare with all the accuracy score in project in fig 2

Fig 5

The accuracy score of the rainstorm is achieved by decision tree algorithmwhich is in fig 3.

Fig 6

The fig 4 explains the prepossessing of the data using polynomial regression.

Output Snapshots:
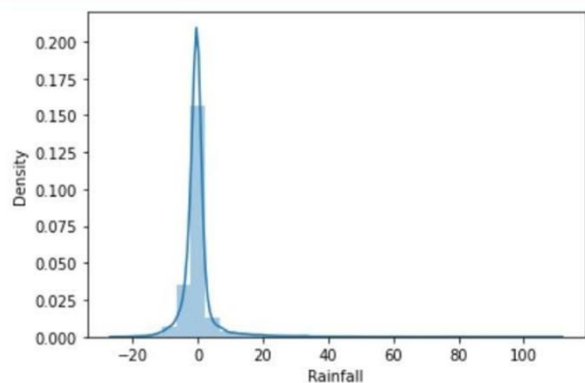
Fig 7: Density with rainfall
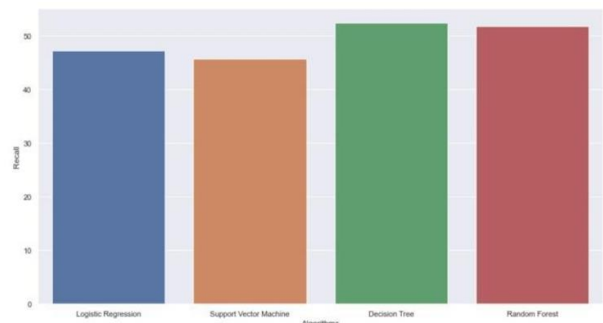
Fig 8:Recall of algorithms
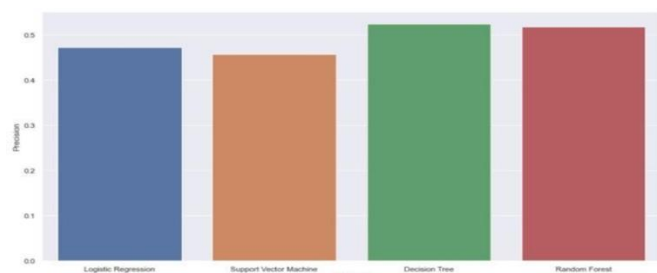
Fig 9: Accuracy of algorithms
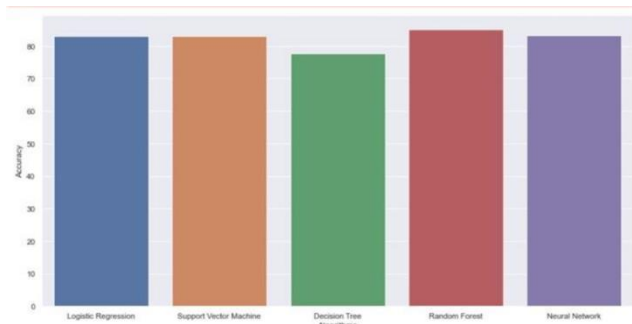
Fig 10 :Precision of algorithms
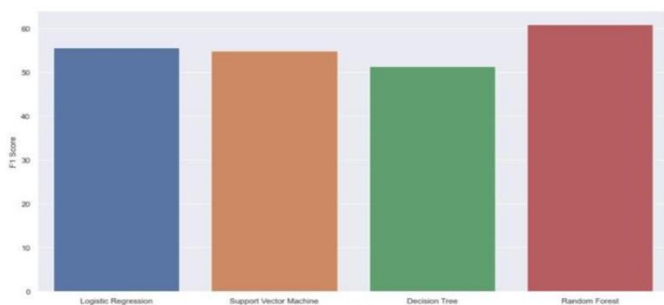
Fig 11: F1 score of algorithms

VIII.   PROPOSED METHODOLOGY

This system proposes an advanced answer for the weather monitoring system using IoT to make its real time data easily accessible over a very wide range. Predicting rainfall is one of the tougher task but using appropriate parameters and classifying themcan help in predicting rainfall.

Classifying the data using SVM algorithm makes it easier to predict the rainfall. The system deals with monitoring weather and climate changes like Temperature, Humidity, Wind speed, Carbon monoxide level in the air, Light intensity, UV radiations, Soil moisture and after analysing and classifying the parameters, rainfall percentage is predicted. The result will be received in the form of graphs and excel sheets. For preprocess , all the result

will be received in the form of different graphs and for machine learning and neural network , the accuracy will be received in the form of Metrics as well as excel sheet and accordingly the predicted value will be received in the form of excel sheet which will contain two columns ID and predicted value. IDs will be same as that of in the datasheet. To get for which region prediction is being done, IDs should be matched with the IDs present in dataset.

A. Ardiuno Uno

It is an open-source physical computing stage based on a easy micro-controller board, and a development environment for writing software for the board. Arduino can be used to develop interactive objects, taking inputs from a variety of switches and or sensors, controlling a variety of lights, motors, and other physical outputs.

Arduino is a single-board microcontroller, intended to make the application of interactive objects or environments more accessible.

Fig 12: Ardiuno Uno Board

B. Cloud Platform (WEB APP)

Cloud platform is designed to store and process Internet of Things (IoT) data. This platform is built to take massive volumes of data generated by devices, sensors, applications, websites and initiate actions for real time responses.

Temperature and humidity sensor The DHT11 is a fundamental, ultra-low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor

and a thermistor to calculate the neighboring air, and spits out a digital signal on the data pin (no analog input pins needed)

IX. CONCLUSION

The conclusion is drawn that a very reliable, efficient and less cost product has been developed which can make life more comfortable and securable. This project does not require any hard installations and can be easily installed in old installations. So, it is easily compatible with old systems. Since the project is prediction based and thus doesn't require any extra cost of installing software. Our project will prove to be efficient for many weather monitoring stations, as it will help them in having an alternative solution when there is any absence of the high cost weather monitoring setup. The rainfall prediction is done with the use of machine learning in minimal costs. The complete weather forecasting setup is flexible enough to be installed anywhere and make weather predictions without much historical experience.

We used different machine learning algorithm to check the accuracy of rainfall prediction. We have compared SVM, Random Forest, Navie Bayes and MLP (Multilayer perceptron) classifiers. From the above figure 3 we can conclude that Random forest is the Machine learning algorithm which is suitable for rainfall prediction in India.

Currently machine learning used in no. of industries. As the data increases the complexity of that data will increase and for that we are using machine for the better understanding of that data. In Rainstorm predictions its pretty helpful with good accuracy score and in rainfall also its gives pretty good predictions. In future we are planning to increase our work in Crop prediction and Flood prediction with the rainstorm prediction

X. REFERENCES

[1]. http://www.arduino.cc

[2]. Data mining techniques for weather prediction: a review, International Journal on Recent and Innovation Trends in Computing and Communication, Volume:2, Issue:8ISSN:2321-8169 2184 – 2189

[3]. Kedarnath flash floods: a hydrological and hydraulic simulation study, Research Communications

[4]. National Remote Sensing Centre, Indian Space Research Organisation, Balanagar, Hyderabad

[5]. Arduino, G., Reggiani, P. and Todini, E., Recent advances in flood forecasting and flood risk assessment. Hydrol. Earth Syst. Sci., 2005, 9(4), 280–284.

[6]. Srinivasan, J., Predicting and managing extreme rainfall. Curr. Sci., 2013, 105(1), 7–8.

[7]. IMD, A preliminary report on heavy rainfall over Uttarakhand during 16–18 June 2013. India Meteorological Department, Ministry of Earth Sciences; July 2013; http://imd.gov.in/doc/uttrakhand_report_04_09_2013.pdf 9 [8]. Auroop R Ganguly, and Karsten Teinhaeuser, ―Data Mining for Climate Change and Impacts,‖ IEEE International Conference on Data Mining, 2008.

[8]. Badhiye S. S., Wakode B. V., Chatur P. N. ―Analysis of Temperature and Humidity Data for Future value prediction‖, IJCSIT Vol. 3 (1), 2012

[9]. http://en.wikipedia.org/wiki/Weather_forecasting#Modern_method

[10]. V. K. Dabhi and S. Chaudhary, ―Hybrid Wavelet-Postfix- GP model for rainfall prediction of Anand region of India,‖ Advances in Artificial Intell., pp. 1-11, 2014.

[11]. Pinky Saikia Dutta, Hitesh Tahbilder, ―Prediction Of Rainfall Using Data mining Technique Over Assam‖, Indian Journal of Computer Science and Engineering (IJCSE),

Quick Peek

Dinesh G.[1], Argha Roy[2], Anshal Chauhan[2], Balantheran Thoranraj[2]

[1]Senior Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]B.E Student, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Security and surveillance are the need of current time as many of us stay outdoors for many purposes. So, to take care and monitoring our properties at real time needs some technology. So, to overcome this problem we are developing a real time surveillance system with the help of facial recognition. To enable the owner of the house with the real time situation we are trying to develop a web application for the owner. All existing systems provide the means of only monitoring the premises the proposed system should be able to identify the faces and activities. It should also serve the purpose of monitoring the other members and the daily activity at the property premises while the owner is away.

Keywords : Facial Recognition, Deep learning, Security, Surveillance, Real time.

I. INTRODUCTION

On a similar plane, evolution of the digital technology has been transformed almost every aspect of our lives, from the work environment and to the home, from professional and to personal. These days, nothing seems unblemished by this trend. Looking at this one industry in particular, home security has developed its own technologically driven transformation. While retina scanners are not that widely used at home, some of the science-fiction-like qualities are rapidly becoming norm. In short, those who are looking at the home security technology in the present will be very much surprised to learn that this is not some old alarm system.

Some of the forecasts reckon that the home security market as a whole will be worth $47 billion by the year of 2020, with the DIY home security system market worth $1.5 billion. While that the market for home security is expanding, especially with cable TV and Internet service providers are introducing their own security system and home automation products, it is a fragmented market, with "a massive array of large and small rivals." The United State market leader is ADT with more than 60 lakhs households subscribed.

As per the recent statistics that the FBI has reported that 1.7 million homes were burglarized in 2014. The same issued report shows that an estimated loss of $3.9 billion was suffered by the victims in the same year. Overall, when the average value was applied to that of the estimated number of burglaries happened, the average dollar loss per burglary offense was $2,251. Roughly, there are 2.5 million burglaries a year with 66% being home break-ins. Police typically solves

only very less number i.e. 13% of total break-in crimes due to lack of information and witnesses.

As per the reports 53% of break-ins occurs during the day time and 47% occur at night time, and it has also been noticed that break-ins are 6% more likely to occur between 6am at morning and 6pm in the evening while most of the members are out of the house.

With only 17% of homes being equipped with home security systems, many burglars are attracted to homes that lack a system.

The study at UNC Charlotte concluded that burglars are commonly males and aged under 25. 12% of burglars admitted to planning their break-in in advance while 41% said it was an impulsive decision. Also, many burglars admitted they had considered factors like proximity to traffic. Approximately 83% of burglars said they look for evidence of an alarm system, and 60% said they do not break-in if they see one is installed.

II. REQUIREMENTS

The requirements for developing a security and surveillance system is high. The system includes Raspberry Pi Board and Camera module. The project needs a good laptop with at least 8gb of ram and software like Visual Studio Code for the development of the website. Python must be installed in system for the deep learning module.

Face recognition, as one of the most successful applications of image analysis, has recently gained very huge attention. It is due to the availability of feasible technologies, including mobile technology. The face recognition processing, which includes major components like face detection, tracking, alignment and also feature extraction, and it points out the technical challenges of building a face recognition system. Face detection performance is going to be key issue, so techniques for dealing with the non-frontal face detection are needed.

Web programming, is also known as web development, it is nothing but creation of dynamic web applications. Web development is divided into 2 types – front-end development (which is also called client-side development) and back-end development (which is also called server-side development). Web push notifications are the notifications that can be sent to user through desktop web and mobile web. All that's required is to send web push notifications is website that actually has web push code installed in it. Which means that brands that don't have apps can invest many of the benefits of push notifications even if they do not have an app.

NoSQL database includes simplicity of designing, simpler horizontal scaling to the clusters of machines and finer control over the availability. The data structures used in NoSQL databases are different from those used by default in relational database which makes some operations much more faster in NoSQL. NoSQL databases offers a concept of the eventual consistency in which database changes are propagated to all nodes so queries for data might not return updated data or might result in reading data that's not accurate which is a problem known as the stale reads.

Raspberry Pi is name of a series of single-board computers made by Raspberry Pi Foundation, a UK charity which aims to educate people in computing and create easier access to education. The purpose of this interface was to standardise that attachment of the cameras modules to processors for the mobile phone industry. The CSI-2 versions of the interface was extremely popular and used on the almost all the mobile phones and devices currently found. With the increasing camera resolution, bandwidth of data transferring from the camera to the processor increases. The CSI-2 specification developed by MIPI Alliance solves the number of problems that was arose when large amounts of data require transfer to the processor.

Node.js is a worker side JavaScript run-time climate. It's open-source, including Google's V8 motor, library for cross-stage similarity, and a center library. Node.js

is basically utilized for non-obstructing, occasion driven workers, because of its Single-strung nature. It's utilized for customary sites and back-end API administrations, however was planned with continuous, push-based models as a top priority. Beside being successful at what it does, Node.js is mainstream since it has an enormous, dynamic, open-source, JavaScript-based biological system. Likewise, it doesn't will in general break similarity between adaptations significantly.

Python is a deciphered, elevated level and universally useful programming language. Python's plan reasoning accentuates code lucidness with its eminent utilization of huge whitespace. Python language is unfathomably simple to utilize and learn for new amateurs and newcomers. The python language is one of the most open programming dialects accessible on the grounds that it has disentangled linguistic structure and not muddled,

which gives more accentuation on common language. Because of its simplicity of learning and utilization, python codes can be handily composed and executed a lot quicker than other programming dialects.

III. SETTING UP YOUR WORK ENVIRONMENT

Facial recognition: Deep learning module should be able to identify the unauthorized persons properly.

Real time database: Real time database is used to store the video content as well as images for retrieval.

Push notification: Push notifications are used to alert the owner in case of any emergency situations.

Capturing real time video feed: Video feed should be captured at real time, so raspberry pi CSI-2 has to be used for high quality video feed.

Computation and receiving video feed: computational works, processing images and videos, working on deep learning module we need computation unit raspberry pi module 4 will take care of it.

Network connection: Network connection should be proper so that video feed/ images gets stored in database. And also user gets notifications.

IV. BRIEF SUMMARY OF THE PROJECT

The owner needs to login in to the system by registering on the network and create an account of him then he is asked for the authorized people picture and the websites this pictures to train the model with the data . and then the camera module (Raspberry pi CSI-2) is installed in Raspberry pi 4 board where it is used to record the live video feed and process it , to be specific the facial recognition is started on then when ever a mismatch occurs the system takes a screenshot and stores in the NoSQL database, and a alert is sent to the user in the mode of a web push notification , where the user can easily Log in by already created account and see the snap shots and even can watch live streams in the mean time the whole camera and other recognition module works simultaneously round the clock.

V. APPLYING SCRUM TO THE DEVELOPMENT PROCESS

"Agile is the ability to create and respond to change. It is a way of dealing with, and ultimately succeeding in, an uncertain and turbulent environment.

The authors of the Agile Manifesto chose "Agile" as the label for this whole idea because that word represented the adaptiveness and response to change which was so important to their approach."[1]

As referenced in the above paragraph our team realized to create our own working model we will be needed to follow such methodology. As a team and a complete new project we can always try new things and respond to the unexpected and refine it to such extent. For this particular project we followed SCRUM of agile methodology. Where we needed to plan our requirements, Make commitments and alternate day team meetings where we have discussed

the progress of our project and did additional refinements to make the project work. Below was our thought process during the entire duration of the project.

- Figure out how to detect a face in the camera so we started working on it and we figured out about openCV.
- Next was to identify what kind of algorithm be used to make the model efficiently working. There we tried two different libraries where the first one ended up consuming the entire RAM on the device as the model has to be deployed on a limited constraint device search for a new algorithm started.
- We found LBPH algorithm was our go to, so we created our model using LBPH algorithm and fitted our own data into it.
- The next part to work on was getting data automatically of the user so we used some openCV features to collect the data using haar based cascade classifier [2].
- Now that we were able to create a working model and an automatic frontal face picture collector, it was time to develop the kind of data to be stored, and we decided to create such a way where only the unauthorized person activity will be stored as an video.
- To store the video while the model is working we optimized the approach where the device doesn't take RAM more than 500MB.
- Now the next part was obviously we can not expect an user to run programs the way a developer does, so we designed a website which eases the user experience and hosted it from a staging server which keeps it running for 24*7.
- Next was a place to store the data for that we used MongoDB and stored the data as user objects where also the video and log files was to be stored for every individual user.
- The last part to develop the real-time video feed access which was created using WebRTC and socket Programming where the user can contact his model sitting at home to access the video feed at any time which keeps listening to a port as we have configured.
- And last step was to deploying the model on Raspberry Pi 4 which we have mentioned and decided to do also. So, we took help of the documentations [3][4] to deploy our models on the Raspberry Pi .
- And with this step we had successfully developed a smart real-time security camera or home security system what we promised to within budget though it can be refined further by modifying certain functions.

VI. CONCLUSION

In case of real time projects, it can be explored more and more as per the scenario requirements. For example, if a new sensor like detecting someone on the door or in the surrounding of the house has to be added, new libraries, hardware and suitable programming language can be used to achieve the goal. When we consider this project from a university level, the access to data is usually limited. In order to improve the query processing, we require huge amount of data that can be in the form of datasets or retrieval from a database.

There are many evolving technologies like blockchain, artificial intelligence, deep

learning. Which can also be used in future based on the ease of access to the libraries and respective programming languages.

Another future work reserved is the addition of the alarm in the home security system which will alert the neighbor's in time of such incident like theft or robbery.

The home security system can further be linked with the home automation system to increase the scope of the system.

Real time surveillance can be improved by adding high definition camera and several other modules or sensors to support.

In this Project "Quick Peek" we implemented Home Security System comprising of notification alerts in the user's laptop or smartphone. The hardware equipment is tested and result is obtained. This project is cost effective. Implementation of this project in present day will effectively provide a peace of mind to the residents. This project can be implemented both in apartments and General stores as well.

The project help us understand the key concepts of microcontrollers and programming languages like html, CSS and JavaScript for the web development part and python, deep learning concepts for the face detection algorithm.

Home security is a rapidly growing field and there are new and improved systems coming out every day. This project will also help us understand the market trend and the level of the product expected in market and how it is better than the other options. It is turning out to be a field of scope and new changes can be made to make it more efficient. It has years to come, because security is of prime importance to one and all.

VII. REFERENCES

[1]. https://en.wikipedia.org/wiki/Home_security

[2]. https://www.agilealliance.org/agile101/

[3]. https://docs.opencv.org/3.4/d2/d99/tutorial_js_face_detection.html

[4]. https://www.pyimagesearch.com/2015/03/30/accessing-the-raspberry-pi-camera-withopencv-and-python/

[5]. https://projects.raspberrypi.org/en/projects/getting-started-with-picamer

Travel Explorer

Sani Hemanth[1], Meghana Kancherla[1], Shiva Sai Reddy[1], Vaishali M Deshmukh[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Rd, near Marathalli, Kaverappa Layout, Kadubeesanahalli, Kadabeesanahalli, Bengaluru, Karnataka, India

ABSTRACT

Travel (also known as voyage – derived from Latin) search engine is the new travel searching technology which is available on the World Wide Web (which is known as www or internet). This also gives new options such as a weather report of seven days so that they can know the status of a place plan accordingly for their desired trip or tour with their family or friends and it also provides a chat window where you can chat with your family and friends and plan the trips accordingly. You can see there are many people using internet, the Internet has become a source for people who wants to search and book their trips or vacations. Our project is mainly created so that it works like a search engine and with different features. The user can get all the search queries related to the search. The recommendation process also happens when the user gives a query, based on the query given by the user the user gets a suggestions of the other users search if they have similar searches. There is speech to text conversion so that it helps the users to speak and get there search results than typing. The search is an intuitive and comely interface for the users to search. It also fetches the result in an ordered way so that the users can browse through it and get what they want were quickly. By using machine learning field, we have developed the project. Crawling is used for get contents in a page, indexing is to arrange and put the contents into the database in order, highlighting for confidence, weather report for knowing the weather and chat to chat with your family and friends.

Keywords— Crawling; Parsing the words; Indexing the pages; Ranking the pages; Auto completion the words; Words get highlighted; Voice are converted to words; Recommendation of sites are given.

I. INTRODUCTION

A search engine is usually used to get any kind of information related to the particular query given by the user to get the desired results or information from database. These search engines are built to be used locally or on a web site and the user can give a query in any form they want such as, html links, words, letters, image for search etc. Therefore, to get a proper result and to achieve contents search engines are used.

Using the whole concept of search engine and getting a result only for A travel search engine is one of the newest travel planning tools which are available on the Internet. As you can see there are many people using internet, the Internet has become a source for people who wants to search and book their trips or vacations.

In these recent days, Travel Search Engine is latest and newest invention in the World Wide Web. This

engine's resulted site doesn't promote any travel based products. The users can enter the details or query related to travel or trip required information and it gets the results by searching various online travel websites in order to get the best possible deal for the user's travel or vacation. Primarily, we can say that a travel search engine does all the work for getting appropriate results for the user. With the use of a travel search engine, one can ease during the planning of a travel or vacation and enjoy their travel or vacation experience to the best level.

There are many benefits in a travel search engine, one of the main benefit is that the user can fetch the details for their travel easily rather than contacting various agents or looking into various travel websites, with this search engine they can get the information much faster. With the use of travel search engine the user can save time, money and energy.

II. RELATED WORK

Search engines: A Comparative Approach - This is the journal written by Silvia Angeloni on December 22, 2020. In this journal the it was explained how search engines are widely used these days. It plays a vital role and it is very helpful for the user to get any related data. Hoe these search engines get the data and display it to the user in a proper format. It makes it easier for the user to get answers to their queries.

An Analysis of Search Engine for Travel - This is a conference paper published by Bing Pan, Zheng Xiang and Rob Law on February 10, 2010. As the technology is growing, travel search engine have become a main part of the internet marketing of tourism business. Online travellers use these search engines for searching about their required travel.

Voyageur: An Experimental Travel search engine - This is a research article published by Sara Evensen, Alon Halevy on May 13, 2019. In this article they described about the Voyageur which is a travel search engine mainly focuses on experiential search on the aspects of the service under consideration. It also

highlights interesting facts and trips about the services.

Trip Advisor - This is an American online travel company which operates as a website and mobile app. This is widely used nowadays. It provides hotels bookings. This is the popular site used these days. Nowadays it's extended to flight travels also.

Internet in Travel and tourism: Expedia - This is a journal of travel published by Rob Law and Freddy Chen. This article said about the growth of Expedia and how it influenced the travellers. This will be useful to potential travellers who plan to make reservations, bookings and plan their travel.

Travelocity becomes a retailer of travel - This is an article is Research Gate published by Barry Smith, Dirk Guenther and Ross Darrow. In this article the market introduction and the growth of Travelocity and how it became famous was explained. At a particular time, this has the highest revenue but it couldn't hold the same for many years. Their contribution was spread worldwide and it became famous.

III. METHODOLOGY

1. The main steps for creating a search engine are:

A. Generating a server for the search engine:

As we know, the search engine requires a huge amount of data and it consumes lot of energy to crawl large data from the internet, hence for this we require a server. Indexing the data and crawling the huge data takes place simultaneously.

For this search engine we should set a search engine and this is tedious job. The reason is that this search engine required a huge amount of resources. So we have to set up a server for the efficient working of the search engine. This sever should have the below three qualities:

- It should sustain high memory load.
- It should have high speed (read/write).
- It should have low network latency.

B. Creating the web crawler for the search engine:

Crawler is one of the main component for the search engine to work efficiently. Web crawler is responsible for fetching data from the internet i.e. the world wide web. When we give one link to the crawler, it digs deep into it and crawls for other links and then from the obtained links again the process is continued. Therefore, it is a recursive process which crawls the internet to obtain the results. It gets the links, title and the short description for each link and adds it into the database.

C. Creating Database:

The information which has been crawled by the program in each link provided to the crawler is then been stored in to the database. When the user types something on the search bar the key word looks for the match in the database to retrieve the information of the keyword. We would need a lot of space for our storage because it should produce a good quality of results to the users. To reduce the delay, we should be using very good performance servers where we can host our search engine.

D. Page Ranking

The page ranking is done when a user gives their query, the query is the key word and the key words targets the words in the database and gets relevant data by fetching from the database. This also helps to fetch unique data from the database for the result. These are also sorted on the base of title, link and description.

E. Building a search user interface

The user communicates with the search engine through the search engine interface to search for their queries and get their desired results. It covers all the complex work at the back end and provides the user with a graphical user which allows the user to interface with the search engine. This graphical interface allows the user to type the query to be searched and then the search engine gets the relevant results and displays it in the interface where the user can view the results obtained with some recommendations given to the user based on their search.

2. System Diagram:
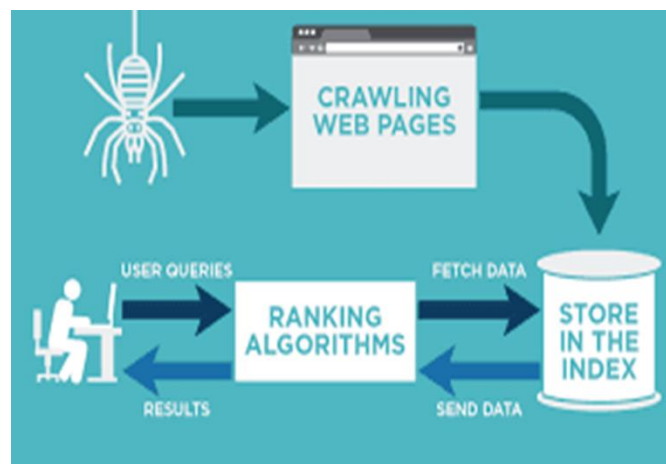
Fig 1: Working of the search engine

3. Architecture of the search engine:
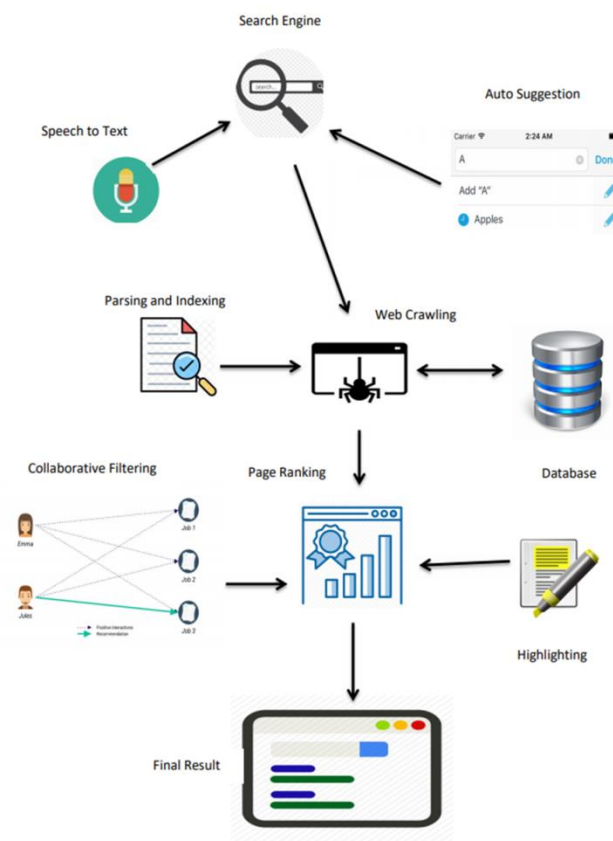
Fig 2: Architecture of search engine

F. Building a weather report:

We get the weather details from OpenWeather API. We have to first get the AP key from it. We have to take the input from the user, for that particular place we find the latitude and longitude. These latitude and longitude are passed to the URL with the API key which gives the weather report in JSON format. This data is retried in php and we display the required details to the user. Seven-day weather report is given to the user which helps the user to plan their travel or trips.

G. Building a chat application:

After setting up the database we should create a login and register for the users to use chat application. Once the user is logged in he can chat with their friends and family. The chat and the list of friends are stored in the database so when they come back all are recorded and they can again plan for different vacation.

IV. RESULT AND DISCUSSION

Travel Explorer is a search engine which is developed using the algorithms mentioned in the methodology. It fetches the data which is the query to be searched. This query can also be taken as a voice message from user and it's converted into text. While the user types the query for each alphabet the user enters auto suggestions are generated to the user, this makes it easier for the user to search for the required results. The process of indexing, parsing and simultaneously crawling processes are performed and the data obtained are stored into the database.

After the query text is taken from the user, the ranking process takes place, which checks for the keyword i.e. the user entered query and it checks the database and displayed the best matched results to the user. In the display screen the title, link and a short description is given. Based on what the current user has searched, recommendations are suggested to that user. The similar user with similar interests as the current user are found and then using the user base filtering algorithm the recommendations are given to the user. This might be helpful for the user to explore these sites which may help the user to book and plan their travel. Based on the place which the user has typed in search, the weather report for seven days is displayed to the user. This might help the user to book and plan their travel or trips accordingly.

The group chat option is available which helps the user to chat whit their family or friends and book their travel. To use this group chat feature the user has to login or register. During register the user has to mention the name, email and password, these values are stored in the database. During login the user is authenticated and then the chat option is available to the user. Finally, this search engine helps the user to get results and plan or book their travel or trips in easy steps.
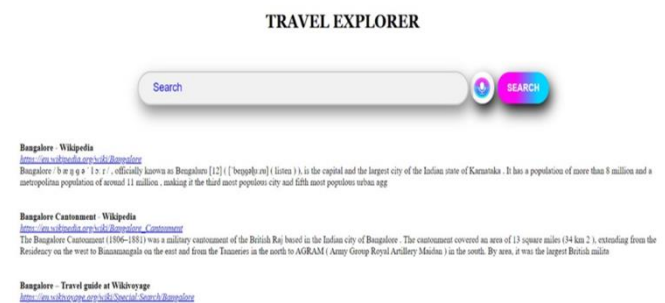
Fig 3: Screenshot of Travel explorer when search text is Bangalore

V. CONCLUSIONS

As we can see, this project is mainly meant to help the user for easy booking of their travel or trips and even more. It serves the purpose of allowing the user to save time by providing results of what the user needs. This search engine gives appropriate and accurate results for the user. The title, link and descriptions are displayed to the user based on the user's search. In the result page the recommendation is given to the user. These recommendations might give additional suggestions to the user during the search. This recommendation system is created using

collaborative filtering algorithm. We use user based filtering method. The similarity between users are matched with the current user and the sites are suggested to the user.

To create this search engine, firstly the process of parsing, indexing and web crawling are followed simultaneously. Then the results found are put into the database. Then secondly, ranking algorithm is followed where the user query is taken as a keyword and based on that the results are filtered and best found links are displayed to the user.

The user can also search using their voice. This voice is taken and converted into text and this text is further sent for processing. To convert the voice to text we use speech - to - text conversion, which is a natural language processing algorithm. When the user types a query based on each letter the user types, a list of auto completion words or sentences are given to the user which makes the search easier for the user. Finally, there is group chat feature available for the user, which helps the user to communicate with family or friends. The user can easily communicate with them and discuss the plan. To use the group chat, the user has to login or register. The details are stored in database. For the user typed place, the weather report is generated for the next seven days from the current day. This weather details might be helpful for the user for booking and planning their travel or trips.

VI. REFERENCES

[1]. Existing Travel search engines are found in : https:/www.search-engine-watch.com/2005/09/14/travel-search-engines

[2]. Steps to create a search engine : https:/azati.ai/build-an-intelligent-search-engine/

[3]. Working process of a search engine : https://moz.com/beginners-guide-to-seo/

[4]. Recommendation system - Collaborative filtering steps : https:/realpython.com/build-recommendation-engine-collaborative-filtering/

[5]. Search engine working and steps to use : https://www.nibusinessinfo.co.uk/content/search-engined-how-do-they-work

[6]. Top search engines based on travel available are : https://www.valuepenguin.com/travel/search-engines

[7]. Top journals and articles on search engine : https:/www.searchenginejournal.com

[8]. "The importance of search engine" - By Gregory Smyth written in inetAsia solutions article.

[9]. The best search engines : https://www.lowcountrygradcenter.org/the-6-best-search-engines

[10]. Trip adviser website which has a travel search engine : https:/www.tripadvisor.co.in/

[11]. Travelocity The travel site with best packagesfor booking : https:/www.smartertravel.com/providers

Emojify Yourself

Surya K[1], S. Sridevi[1*]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The project presents an approach for image cartoonization. This could be often supported by the observation of the behaviour of cartoon painting and after the project had been consulted by the artists, it's been projected to spot on an individual basis the 3 white-box illustrations from the pictures, the surface illustration that contains a sleek surface of cartoon pictures, the structure illustration that refers to the distributed colour-blocks and flatten universal content among the celluloid vogue advancement, and so the texture illustration that reflects the high-frequency texture, contours, and details in cartoon pictures. Within the work, the project was composed of a try of objects throughout a conditional GAN framework using a completely unique consistent composition-by decomposition network. Indicated by the item pictures from 2 distinct distributions, the model had generated a wise composite image from the joint distribution following the feel and form of the input objects.

Keywords— Image cartoonization, sparse colour-blocks, high-frequency texture, contours, conditional GAN framework

I. INTRODUCTION

Cartoons are commonly used in various kinds of applications. They are artistically made and they require elegant and fine human artistic skills. While portraying the cartoons in humongous numbers, implementing this technique was time consuming for the artist as it needs to define the sketch of the particular cartoon properly to get a good result.

To overcome the problem faced by the artist in this project, the images have been created with the help of GAN in which they are converted. A couple of years ago, the styling of images had consisted of a specific domain named "non-photorealistic rendering" had been developed.

The traditional algorithm was developed on the bottom of the domain for the styling of images and that it was successful in styling any image by adding the designs, texture, effects, etc in this project. With the assistance of the algorithm, many softwares were developed to convert the real images(snapshots) into cartoon images.

II. LITERATURE SURVEY

The authors studied various problems faced by the sketch artist while sketching the various black and white cartoon drawings, for the consideration of colouring of the various sketches, mixing of different colours to get a unique shade for a particular sketch.

According to the research, some difficulties were faced by the artist to get a unique or a desired colour it needs after mixing two or more colors. So to overcome the problem, an application was introduced which was known as "drawing-to-picture synthesis while utilising conditional generative adversarial networks" (cGAN).

Later on, the project itself had found that the application faced a problem and had failed to give the desired output. To avoid the issue, the project itself had invented the self-interior design model which could automatically create a suitable hue for a sketch.

The application was constructed on the conditional GAN with 'Unet'structure which allowed the output image to have either the subordinate information of sketch as well as the learned upper-class colour information.

The project had also found more restrictions based on the pix2pix model to get a finer painting. Here the project had worked on the auto painter to alter to hue command so that the matrix could alter the combined result which could satisfy the user by various colours.

III. PROPOSED SYSTEM

Fig. 1 shows a block diagram of a GAN framework consisting of a generator G and 2 discriminators.

The plan of image cartoonization block diagram framework was given. The area unit of the pictures was rotten into the exterior illustration, the construction illustration, and also the texture representations and the 3 freelance modules area unit were introduced to takeout comparable representations.

A GAN framework with a generator G and a couple of discriminators Ds and Dt were planned, where Ds aims to inform apart between exterior illustration taken out from model outputs and cartoons, and Dt was used to inform apart between the feel illustration extracted from outputs and cartoons. Pre-instructed VGG network was used to get rid of the superior choices and had obligatory the abstraction restriction on international merchandise between the removed structure representations and outputs, and jointly between input photos and yields. Weight for each half was adjusted among the loss operate, that had allowable the users to control the output vogue and custom-made the model to numerous use cases.

Along with the proposed three-step approach, pre-processing was an important part of the model. It had helped to smoothen the image, filter the features, converting it into the sketches, and translating the output from a domain to another domain. After implementing the related work and it was made sure that the output generated by the model will give the project the best output which will retain the highest quality features.

Fig. 2 shows an example of a design diagram(e.g Sequence Diagram)

Sequence Diagram - Object interactions arranged in the time sequence.

Firstly, the user logins into the system software. It verifies whether the person is an authorized user or not. It inputs the image and gives its output image.
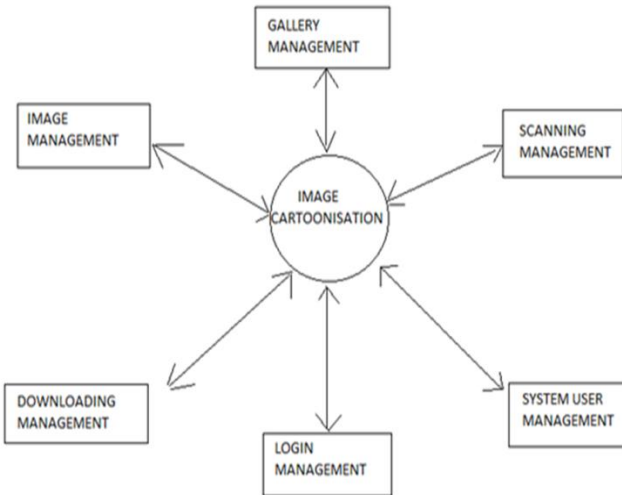
Fig. 3 shows an example of a design diagram(e.g Data Flow Diagram Level-0)

Data Flow Diagram Level-0 :-

Represent the flow of data in a business information system. Image cartoonization's function was to show the process. Image management's function was to show how the images are managed during the cartoonisation. Gallery management's function shows how the images are storeed and made changes.
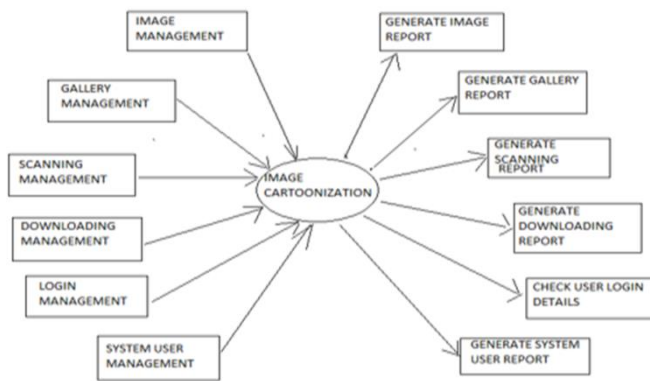
Fig. 4 shows an example of a design diagram(e.g Data Flow Diagram Level-1)

Data Flow Diagram Level-1 - Image report's function was to generate the report.User login details' function was to check the login details of the user.
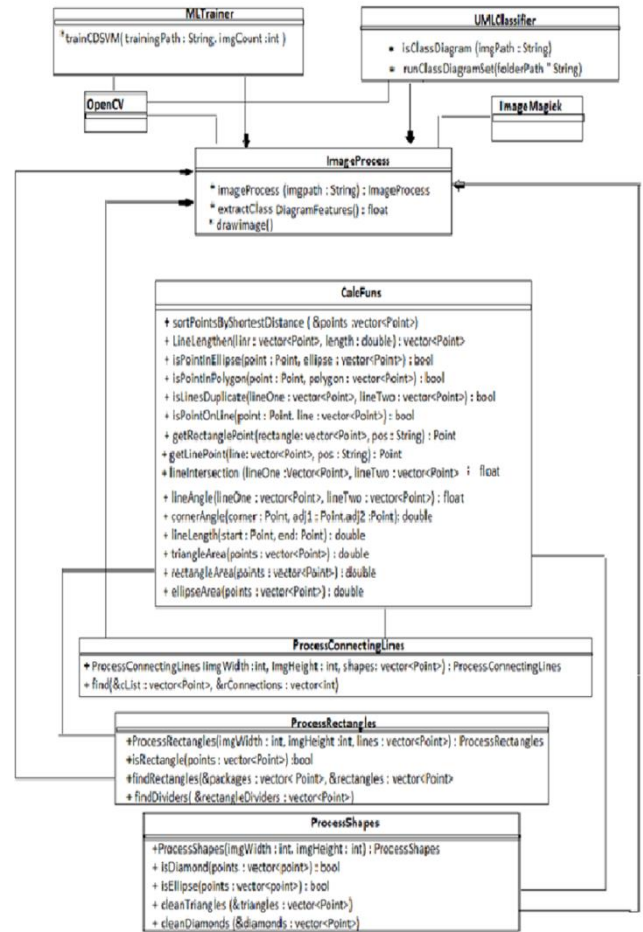
Fig. 5 shows an example of a design diagram(e.g Class Diagram)

Class Diagram:- Translating the models into programming code.
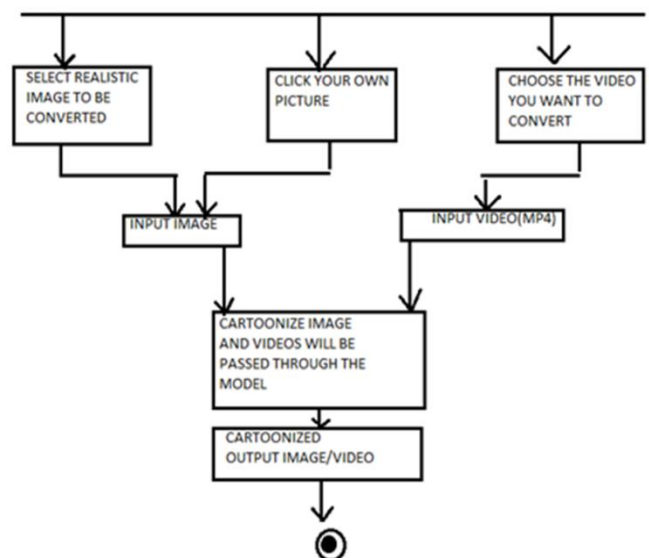
Fig. 6 shows an example of a design diagram (e.g Activity Diagram)

Activity Diagram:-Depicts the behaviour of a system. Select the realistic image to be converted and click its own pictures and input the images. Images will be passed through the model and output image will be cartoonized.

IV. IMPLEMENTATION

Algorithm:-

- Firstly applied the bilateral filter to cut back the colour palette of the image.
- Then converted the particular image to grayscale.
- Now applied the median blur to cut back image noise within the grayscale image.
- Created a mask from the grayscale image using adaptive thresholding.
- Finally mixed the color image produced from one step with edge mask produced from another step.

The project has used the following modules:-

- Numpy: Images are stored and processed as numbers. These are taken as arrays. It had used NumPy to deal with the arrays.
- CV2: It has been imported to use OpenCV for image processing

First the project has downscaled the image and then applied bilateral filter to get a cartoon flavor. Then again the project has upscaled the image.

Next step was getting a blurred version of the original image. Now, it doesn't want the colours to interfere in this process. It only wants the blurring of the boundaries. For this, it first had converted the image to gray–scale and then it had applied the media blur filter.

Next step was to identify the edges in the image and then added the image to the previously modified images to get a sketch pen effect. For this first it had used the adaptive threshold. It had been experimented with other types of threshold techniques also. It was because Computer Vision is all about experimenting. The project has compiled the final images obtained from the steps.

The output will be implemented by the project in the future.

V. CONCLUSIONS

This project will help everyone to understand about cartoonization of images using Machine Learning in the best possible way and also assist them to its practical usage.

VI. REFERENCES

[1]. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis

[2]. 2015 IEEE Conference on pc Vision and Pattern Recognition (CVPR) page 1-9. (June 2015)

[3]. Y. Chen, Y.-K. Lai, Y.-J. Liu, "CartoonGAN: Generative Adversarial Network for photo cartoonization", International Conference on Image Processing, 2018

[4]. J. Bruna, P. Sprechmann, and Y. LeCun., "Superresolution with deep convolutional sufficient statistics" In International Conference on Learning Representations (ICLR), 2016.

[5]. K. Beaulieu and D. Dalisay, "Machine Learning Mastery", Machine Learning Mastery,2019.

[6]. J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In Advances in Neural Information Processing Systems, pages 82–90, 2016.

[7]. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via L0 gradient minimization. ACM Transactions on Graphics, 30(6):174, 2011.

[8]. M. Yang, S. Lin, P. Luo, L. Lin, and H. Chao. Semantics driven portrait cartoon stylization. In

International Conference on Image Processing, 2010.

[9]. He Huang, Philip S. Yu and Changhu Wang," An Introduction to Image Synthesis with Generative Adversarial Nets", 2018.

[10]. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In International Conference on Computer Vision, 2017.

Laboratory Testing Strategy Recommender

Dr. Senthil Kumar R[1], Punyasri H[1], Konda Charitha[1], Raju Harshavardhan Reddy[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore,
Karnataka, India

ABSTRACT

The paper is about an application which does the analysis which has to be done by doctor so that it avoids patient consulting doctor before getting test done. In this application user/patient can specify symptoms of covid-19 and based on the symptoms specified the severity of covid-19(mild, moderate, severe) will be detected and the laboratory tests which they have to undergo will be recommended and details of the test. And the best laboratories nearby his/her location will be recommended. This reduces the time consumption of the patient visiting the doctor for symptom analysis. Decision Tree Classifier is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. It helps to think about all the possible outcomes for a problem. There is less requirement of data cleaning compared to other algorithms. Present application aims in helping the user in way that, the user can specify symptoms of covid-19 and according to the symptoms specified the tests which they have to undergo will be recommended. It also aims in recommending the details of the laboratory tests the patient undergoes. It recommends the nearby laboratory to the patient/customer.

Keywords — Covid-19, Decision Tree Classifier, Severity, Accuracy, Imbalanced dataset, Correlation matrix, Label encoder.

I. INTRODUCTION

Now a day's people are more interested in new technologies which is saving most of time. Even though if a person is suffering from any illness, he/she need to consult doctor and then according to doctor recommendation the patient need to get the tests done and he have to revisit doctor for analysis of the reports which is time consuming. The proposed system will be an application helpful to the user in way that, user/patient can specify symptoms of covid-19 and based on the symptoms specified the severity

of covid-19(mild, moderate, severe) will be detected and the laboratory tests which they have to undergo will be recommended and details of the test. And the best laboratories nearby his/her location will be recommended. This reduces the time consumption of the patient visiting the doctor for symptom analysis. Decision Tree Classifier is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and

each leaf node represents the outcome. It helps to think about all the possible outcomes for a problem. There is less requirement of data cleaning compared to other algorithms.

II. LITERATURE SURVEY

Classification algorithms can be applied on unstructured or structured data. Classification is a technique where data is categorized into number of classes. Identifying the class to which a new data will fall under is the main aim of classification problem.

A. Classification

There are different types of classification methods.

(1) Logistic regression:

Calculated Regression is an AI order calculation i.e., CLASSIFIER. Produces brings about a twofold organization. So the results are discrete/absolute, for example, 0 or 1, yes or no, valid or bogus, high and low. Calculated relapse is intended for this reason (characterization), and is generally valuable for understanding the impact of a few autonomous factors on a solitary result variable. Works just when the anticipated variable is double, accepts all indicators are autonomous of one another, and expects information is liberated from missing qualities. (2) Naïve Bayes: Naive Bayes calculation depends on Bayes' hypothesis. With the supposition of autonomy between each pair of highlights Naive Bayes classifier function admirably in some certifiable circumstances i.e., spam sifting and report arrangement. These are quick contrasted with more modern techniques. It requires limited quantity of preparing information to gauge the fundamental ascribes. The primary drawback of Naive Bayes is it is known to be an awful assessor. (3) Stochastic angle plummet: It is a basic and effective strategy to fit straight models. At the point when the quantity of tests is exceptionally enormous it is explicitly valuable. It upholds diverse misfortune capacities and punishments for characterization. The upsides of this calculation are

productivity and simplicity of execution. The disservices are it's anything but various hyper-boundaries and it is delicate to include scaling. (4) K-Nearest Neighbors: Neighbors is a languid learning. It is alleged in light of the fact that it essentially stores cases of the preparation information. What's more, doesn't endeavor to build an overall interior model. From a basic greater part vote of the k closest neighbors of each point order is processed. This calculation is easy to execute, strong to boisterous preparing information, and viable if preparing information is huge. The impediments are need to decide the worth of K. Since it needs to process distance of each example to all the preparation tests cost of calculation is high.(5)Decision Tree: Decision tree is easy to comprehend. It is additionally simple to picture. Choice tree requires little information . The disservice of this calculation is tree can make intricacy. Made tree can be unsteady in view of little varieties in the information. This can prompt the production of totally unique tree once in a while. (6)Random woodland: Random timberland classifier is named as meta-assessor. It is alleged on the grounds that it's anything but various choice trees on various sub examples of datasets. It utilizes normal to improve the precision expectation of model. Also, henceforth powers over fitting. The size of sub example is same as the size of unique info consistently. In any case, the examples are drawn with substitution. As far as over fitting Random backwoods beat choice trees in a large portion of the cases. This classifier has moderate constant expectation. It is hard to execute and intricacy of calculation is high.(7) Support vector machine: SVM is an addresses the preparation information as focuses in space isolated into classifications by a reasonable hole that is pretty much as wide as could really be expected. Further new models are planned into that equivalent space and afterward anticipated to have a place with a classification dependent on side of the hole they fall. This classifier is better in high dimensional spaces. And afterward it's anything but a subset of preparing

focuses in the choice capacity so it is additionally memory productive. The calculation doesn't straightforwardly give likelihood appraises, these are determined utilizing a costly five-overlap cross-approval.

III. METHODOLOGY

A. Multi-Class Classification

Grouping issues having various classes with imbalanced dataset present an unexpected test in comparison to a twofold characterization issue. The slanted dissemination makes numerous regular AI calculations less successful, particularly in anticipating minority class models. To do as such, let us initially comprehend the current issue and afterward talk about the approaches to defeat those. Multiclass Classification: A grouping task with multiple classes; e.g., order a bunch of pictures of natural products which might be oranges, apples, or pears. Multi-class characterization makes the supposition that each example is doled out to one and only one name: a natural product can be either an apple or a pear yet not both simultaneously. Imbalanced Dataset: Imbalanced information regularly alludes to an issue with characterization issues where the classes are not addressed similarly model, in datasets like those that describe false exchanges are imbalanced. By far most of the exchanges will be in the "Not-Fraud" class and a little minority will be in the "Misrepresentation" class.

B. Decision Tree Classifier

A choice tree is a flowchart-like tree structure where an inside hub addresses highlight (or trait), the branch addresses a choice principle, and each leaf hub addresses the result. The highest hub in a choice tree is known as the root hub. It figures out how to parcel based on the quality worth. It segments the tree in recursively way call recursive parceling. This flowchart-like construction helps you in dynamic. It's anything but's a flowchart graph which effectively

copies the human level reasoning. That is the reason choice trees are straightforward and decipher.
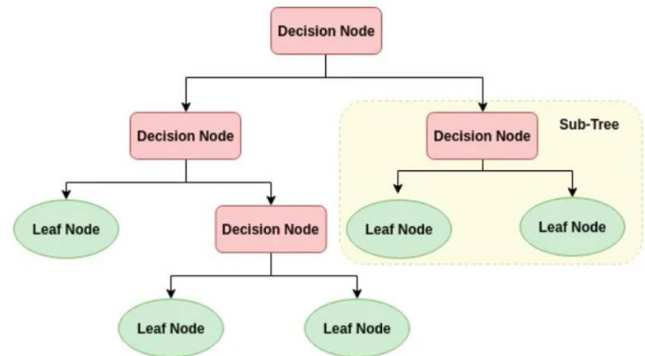
Fig. 1 Basic Decision Tree

IMPLEMENTATION STEPS FOR DECISION TREE CLASSIFIER:

1) Importing required libraries- first load the required libraries.
2) Loading Data- load the required Covid-19 symptoms dataset using pandas' read CSV function.
3) Feature Selection- need to divide given columns into two types of variables dependent (or target variable) and independent variable (or feature variables).
4) Splitting Data- need to split the data into training data and test data using train_test_split(), need to pass 3 parameters features, target, and test_set size.
5) Building Decision Tree Model- create decision tree model using DecisionTreeClassifier().
6) Evaluation of Model- Accuracy can be computed by comparing actual test set values and predicted values.
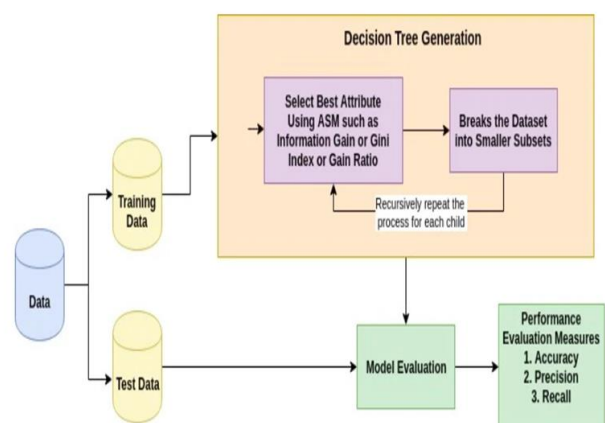
Fig. 2 Steps of Implementation of Decision Tree
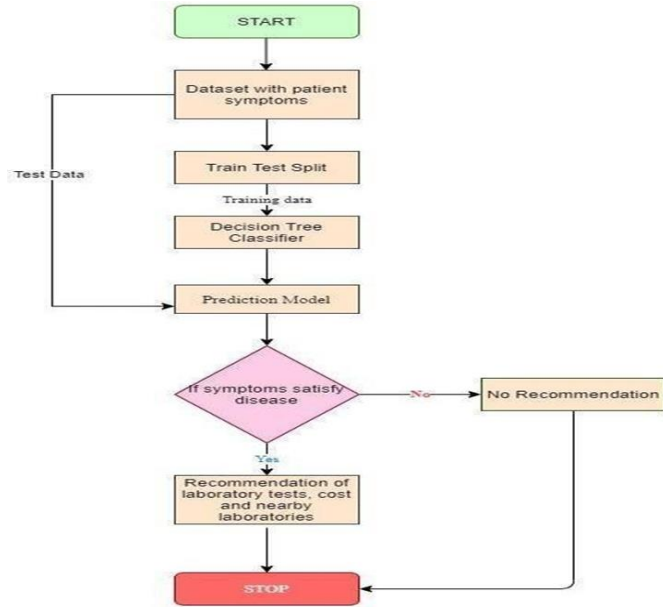
Fig. 3 Flow Diagram for Laboratory Testing Strategy Recommender

DETAILS OF THE DATASET:

Dataset name- Covid-19-severity-classification

Attributes or predictors or Features:

1) Difficulty-in-Breathing
2) Fever
3) Tiredness
4) Dry-Cough
5) Sore-Throat
6) Body-Pains
7) Nasal-Congestion
8) Runny-Nose
9) Diarrhea
10) Loss-of-taste-smell

Target variable:

    SEVERITY

Target variable is a multi-class variable where severity is divided into three classes:

1) MILD
2) MODERATE
3) SEVERE

The Dataset designed has 11 columns where 10 columns are predictors and 1 column is target and 500 rows.

C. Results

In this paper, there are 10 side effects of Coronavirus which are free factors and 1 objective trait for example seriousness of Coronavirus which is reliant variable. Target is a multi-class variable where seriousness of Coronavirus is separated into 3 multi-classes Mild, Moderate and Severe. First the necessary modules are imported and afterward information is isolated into indicators and target. After detachment the information is parted into preparing and test information. In the wake of parting DecisionTreeClassifier () object is made and afterward the train information is fitted into the model. Utilizing test information model forecast is done and model assessment is finished by figuring the exactness of the model.

Fig. 4 Covid-19-severity-classification dataset info

1) Import Required Libraries and Load Dataset:

Fig. 5 Loading Dataset

2) Label Encoder pre-processing technique:

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['Difficulty-in-Breathing']=le.fit_transform(df['Difficulty-in-Breathing'])
df['Fever']=le.fit_transform(df['Fever'])
df['Tiredness']=le.fit_transform(df['Tiredness'])
df['Dry-Cough']=le.fit_transform(df['Dry-Cough'])
df['Sore-Throat']=le.fit_transform(df['Sore-Throat'])
df['Body-Pains']=le.fit_transform(df['Body-Pains'])
df['Nasal-Congestion']=le.fit_transform(df['Nasal-Congestion'])
df['Runny-Nose']=le.fit_transform(df['Runny-Nose'])
df['Diarrhea']=le.fit_transform(df['Diarrhea'])
df['Loss-of-taste-smell']=le.fit_transform(df['Loss-of-taste-smell'])

df['SEVERITY']=le.fit_transform(df['SEVERITY'])
df.head()
```

	Difficulty-in-Breathing	Fever	Tiredness	Dry-Cough	Sore-Throat	Body-Pains	Nasal-Congestion	Runny-Nose	Diarrhea	Loss-of-taste-smell	SEVERITY
0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	1	1	1
3	0	0	0	0	0	0	0	1	0	0	0
4	0	0	0	0	0	0	0	1	0	1	1

Fig. 6 Label Encoder

3) Spilt the dataset into training and test data and build prediction model using Decision Tree Classifier:

```
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1) # 70% training and 30% test
```

```
# Create Decision Tree classifer object
clf = DecisionTreeClassifier()
```

```
# Train Decision Tree Classifer
clf=clf.fit(X_train,y_train)
```

```
#Predict the response for test dataset
y_pred = clf.predict(X_test)
```

```
# Model Accuracy, how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred)*100)
```

Accuracy: 98.88888888888889

Fig. 7 Accuracy

```
corr=df.corr()
fig, ax=plt.subplots(figsize=(8,6))
sns.heatmap(corr, annot=True, ax=ax)
```

<matplotlib.axes._subplots.AxesSubplot at 0x1f449ee4130>
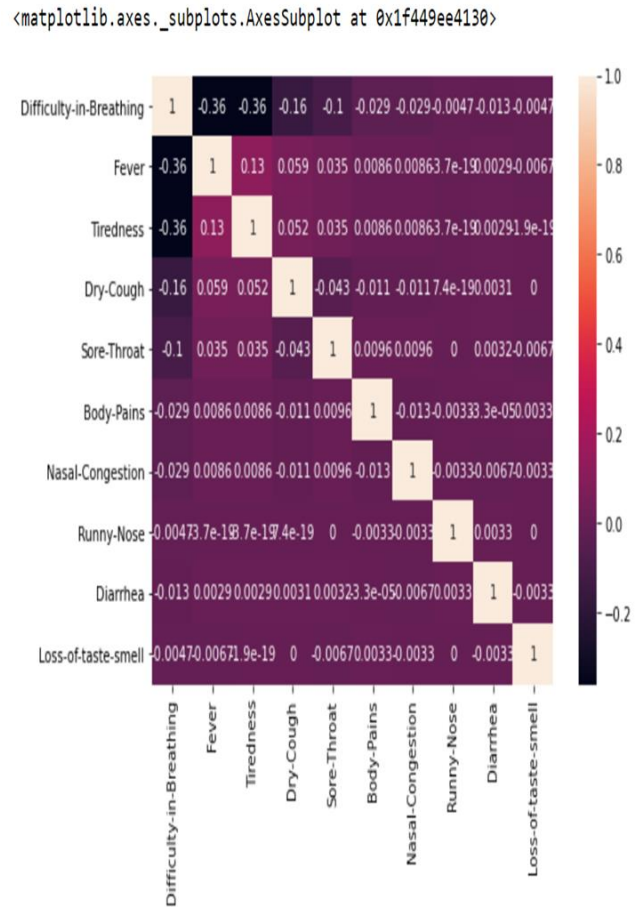
Fig. 8 Correlation Matrix

ACCURACY OF MODEL USING DECISION TREE CLASSIFIER IS 98.88%.

IV. CONCLUSIONS

There has been dramatic development in advanced data and there are huge number of decisions for items and administrations. Despite the fact that if an individual is experiencing any sickness he/she need to counsel specialist and afterward as indicated by specialist proposal the patient need to complete the tests and he need to return to specialist for investigation of the reports which is tedious. Along these lines, there is need to channel, focus on and proficiently convey pertinent data to effectively handle the issue of data over-burden. Recommenders tackle this issue via looking through enormous

volume of powerfully produced data to furnish clients with customized substance and administrations. Present application points in aiding the client such that, the client can indicate manifestations of his/her sickness and as per the side effects determined the tests which they need to go through will be suggested. It likewise points in suggesting the expenses of the research center tests the patient goes through. It prescribes the close by research facility to the patient/client. This targets giving computerized help to the specialists. Recommenders are the one aides in taking care of the issue of data over-burden. This venture with the guide of Decision tree classifier predicts the illness and research center tests in best way dependent on the manifestations of the patient. The objective of this application is to give the client the best research facilities accessible and furthermore prescribe the client with tests to be done dependent on the seriousness of Coronavirus which is recognized by the indications and furthermore give the all out subtleties of the test in various labs accessible.

V. ACKNOWLEDGMENT

VI. REFERENCES

[1]. Ishtiaq Ahmed, Shiyong Lu, Changxin Bai, Fahima Amin Bhuyan," Diagnosis Recommendation Using Machine Learning Scientific Workflow".

[2]. Filippo Amato , Alberto López, Eladia María Peña-Méndez, Petr Vaňhara , Aleš Hampl, Josef Havel," Artificial neural networks in medical diagnosis" ISSN 1214-0287 Received 17th December 2012. Published online 7th January 2013.

[3]. Ayedh abdulaziz Mohsena, Muneer Alsurorib , Buthiena Aldobaid " New Approach to Medical Diagnosis Using Artificial Neural Network and Decision Tree Algorithm: Application to Dental Diseases" Received: 28 September 2018; Accepted: 14 December 2018; Published: 08 July 2019.

[4]. Vanisree K Jyothi Singaraju "Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks".

[5]. Cai, Jie, Jiawei Luo, Shulin Wang, and Sheng Yang. "Feature selection in machine learning: A new perspective." Neurocomputing 300 (2018): 70-79.

[6]. How to perform feature selection (i.e. pick important variables) using Boruta Package in R ? GUEST BLOG, MARCH 22, 2016.

[7]. Heaton, Je. "Feature Importance in Supervised Training." Predictive Analytics and Futurism (2018): 22.

[8]. Feature Selection Using Genetic Algorithms (GA) in R by Pablo Casas Jan. 23, 19.

[9]. Chen, Keqin, Kun Zhu, Yixin Meng, Amit Yadav, and Asif Khan. "Mixed Credit Scoring Model of Logistic Regression and Evidence Weight in the Background of Big Data." In International Conference on Intelligent Systems Design and Applications, pp. 435-443. Springer, Cham, 2018.

[10]. He, Yaoyao, Yang Qin, Shuo Wang, Xu Wang, and Chao Wang. "Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network." Applied energy 233 (2019): 565-575.

[11]. Khedher, Laila, Javier Ramírez, Juan Manuel Górriz, Abdelbasset Brahim, Fermín Segovia, and Alzheimer's Disease Neuroimaging Initiative. "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images." Neurocomputing 151 (2015): 139-150.

[12]. Barbu, Adrian, Yiyuan She, Liangjing Ding, and Gary Gramajo. "Feature selection with annealing for computer vision and big data learning." IEEE transactions on pattern analysis and machine intelligence 39, no. 2 (2016):272-286.

[13]. Hwang, Jing-Shiang, and Tsuey-Hwa Hu. "A stepwise regression algorithm for high-dimensional variable selection." Journal of Statistical Computation and Simulation 85, no. 9 (2015): 1793-1806.

[14]. Emura, Takeshi, Shigeyuki Matsui, and Hsuan-Yu Chen. "compound. Cox: univariate feature selection and compound covariate for predicting survival." Computer methods and programs in biomedicine 168 (2019): 21-37.

[15]. Aich, Satyabrata, Ahmed Abdulhakim Al-Absi, Kueh Lee Hui, John Tark Lee, and Mangal Sain. "A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques." In 2018 20th International Conference on Advanced Communication Technology (ICACT), pp. 139-143. IEEE, 2018.

[16]. Jia, Feng, Yaguo Lei, Liang Guo, Jing Lin, and Saibo Xing. "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines." Neurocomputing 272 (2018): 619-628.

[17]. Mandy Sidana. "Types of classification algorithms in Machine Learning".(2017)

[18]. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[19]. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[20]. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[21]. FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

[22]. "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[23]. A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[24]. J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[25]. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.

Survey on the Applications of Artificial Intelligence in Cyber Security

Shilpashree U[1]

[1]Department of Computer Science and Engineering, the Oxford College of Engineering Bangalore, Karnataka, India

ABSTRACT

Cyber security has become a major concern in the digital era. Data breaches, ID theft, cracking the captcha, and other such stories abound, affecting millions of individuals as well as organizations. Cyber attacks can ruin the reputation of an organization or let down the organization completely. This research examines the use of AI in the enhancement of cyber security. Recent developments in artificial intelligence are transformational and have exceeded the level of human performance in tasks such as data analytics. The study adopted the thematic literature review method, and data were sourced from Google scholar, science direct, research gates, academia, and others. The investigation revealed that application of AI in controlling cyber attack has advantages and disadvantages. In this chapter, we discuss specific techniques in artificial intelligence that are promising. They cover the applications of those techniques in cyber security. End the discussion with the future scope of artificial intelligence and cyber security.

Keywords : Artificial Intelligence, Intelligent Agents, Cyber Security, Neural Nets, Expert Systems, Deep Learning (DL), Machine Learning (ML), Scanning Engine.

I. INTRODUCTION

Protecting the computers and networks from any damage, attack or unauthorized access is called Cyber security. Artificial Intelligence is concerned with building smart machines that can conduct and complete tasks without human interference. Cyber security and Artificial Intelligence can be used together to avoid cybercrimes. The day to day raising and progressing cyber security threat facing global businesses can be reduced by the integration of Artificial Intelligence into cyber security systems. Machine learning and Artificial Intelligence (AI) are being connected more extensively crosswise over industries and applications than any other time in recent memory as computing power, storage capacities and data collection increase. In present days, most organizations are aware of how even a small cyber-attack would create a lot of damage to their data and create a huge amount of losses. Therefore, to provide cyber security, they use different lines of safety-Firewall, Antivirus Software. Using these two, backups are done regularly as a part of the disaster recovery plan. But these two lines of safety require the work of a Professional. However, using Artificial Intelligence and cyber security would change this traditional approach. Using human intelligence to focus on creating Artificial Intelligence algorithms and assigning cyber security responsibilities to Artificial intelligence will provide high-quality results

and a cybercrime safe environment. Security Professionals are using Artificial Intelligence and Machine learning, a subdivision of artificial intelligence to learn about new cybercrimes, pin out and prevent cyber-attacks with minimal human intervention.

II. ARTIFICIAL INTELLIGENCE IN CYBER SECURITY

The concept of AI was proposed in the year 1956 by John McCarthy as the science and engineering of producing intelligent automata, particularly intelligent computer applications. It is concerned with how to make computers think, work, learn and behave intelligently like humans. The application of AI now affects several aspects of human experience such as expert systems, computer vision, pattern recognition, speech recognition, language translation, robotics, biometric systems, and internet of things (IoT) among others. Artificial Intelligence researchers were interested in developing programs to decrease human work, while security professionals trying to fix the outflow of information. But the two fields have grown closer over the time, when the attacks have targeted to simulate the genuine performance, not only at the human user level but also at lower system levels. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a very good example of connection of artificial intelligence and security. This requires end-user to insert the letters of some unfair image, on some occasions with the addition of a masked sequence of letters or digits that appears on the screen. Improvements in automatic character recognition software, which can be considered to be a reasonable advance in AI technology, could motivate the field towards more refined pattern recognition. So in the practice of trying to secure properties, such as online ticket reservations, the profitable security market is in a way stimulating advances in artificial intelligence. Artificial Intelligence helps us in quickly identifying

and analysing new exploits and weaknesses to help ease further attacks and is an integral part of our solutions. Artificial Intelligence practices are the key to Interference detection and make it possible to respond even to anonymous threats before spreading itself. Artificial Intelligence systems that are intended to learn and adapt, and are proficient of identifying even the minutes of changes in the settings, have the potential to act much earlier and based on vast trove of data than humans when it comes to analysing novel types of cyber-attacks.

III. ARTIFICIAL INTELLIGENCE TECHNIQUES FOR CYBER SECURITY

AI can be said to possess some degree of human intelligence: a store of domain-specific knowledge; mechanisms to acquire new knowledge; and mechanisms to put that knowledge to use. Machine learning, expert systems, neural networks, and deep learning are all examples or subsets of AI technology today.

3.1. Expert systems

Are programs designed to solve problems within specialized domains, by mimicking the thinking of human experts, they solve problems and make decisions using fuzzy rules-based reasoning through carefully curate   bodies of knowledge. CSIA - Cyber Security Artificial Intelligence Expert System has knowledge-based systems are composed of two sub-systems: The Knowledge Base and the Inference Engine. The knowledge base represents the illustrations and assertions in the real world. The Inference Engine is an automatic reasoning system. It evaluates the current situation of the knowledge base and applies the rules relevant to that, then asserts new knowledge in to it.

Expert System
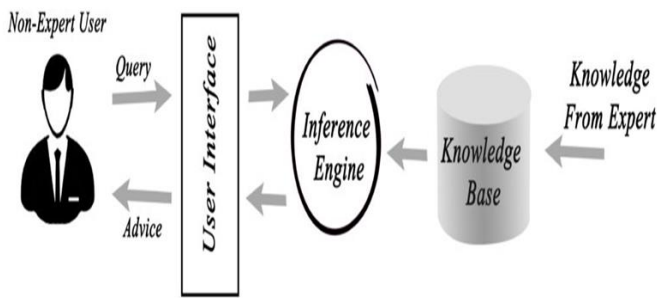
3.2. Neural networks and Deep Learning

The inventor of the first neuro computer, Dr. Robert Hecht-Nielsen, defines a neural network as –"**...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs.**" ANN has the ability to learn and solve problems in different complex domains. It can learn from data in any domain and address absorbing concerns by merging with disparate nerves. In cyber security, ANNs have been used within all four stages of integrated security approach (a holistic categorization of cyber Défense framework), consisting of early warning phase, prevention phase, detection phase and reactive/response phase. ANN can be used to monitor traffic flow in computer networks when integrated in cyber security, thereby detecting malicious intrusions before an actual attack and hindering cyberattacks eventually through perimeter defence. ANN can learn from previous network activities and assaults so as to avoid later attacks. When deep learning (DL) - an advanced form of ANN- is applied to cyber security, the system can recognize suspicious as well as legal files with no human intervention. This method produces better outcome in identifying threats than the conventional methods applied in cyber defence. Frank Rosenblatt created an artificial neuron (Perceptron) which paved the way for neural networks. This perceptron can learn and tackle absorbing issues by combining with other nerves i.e., perceptron. Perceptron learn on their own to identify the entity on which they are trained by learning and processing the high-level raw data, as our brain learns in its own from the raw data using our sensory organ's inputs. When we apply this deep learning (trained) to cyber security, the system can identify whether a file is malicious or legitimate without human interference. This technique yields a strong result in detecting the malicious threats, compared with classical machine learning systems.

3.3. Intelligent agent

An AI system can be defined as the study of the rational agent and its environment. The agents sense the environment through sensors and act on their environment through actuators. An AI agent can have mental properties such as knowledge, belief, intention, etc. Intelligent agents may also learn or use information to achieve their objectives. They may have responsive characteristics, and when communicating with other autonomous agents they may understand and respond to changes in their domain. They can adapt to real time, learn new things rapidly through communication with environment, and have memory based standard storage and recovery abilities. Intelligent agent is created in showdown against Distributed Denial of Service (DDoS) attacks. PEAS is a type of model on which an AI agent works upon P: Performance measure E: Environment A: Actuators S: Sensors. In case if there is any legal or business issue, it should be manageable to develop a "Cyber Police". Cyber Police should have mobile intelligent agents. For this we should device the infrastructure to support the quality and interaction between the intelligent agents. Multi-agent tools will give a lot of full-fledged operative appearance of the cyber police.

IV. APPLICATIONOF ARTIFICIAL INTELLIGENCE IN CYBER SECURITY

4.1. VULNERABLITY MANAGEMENT

Currently, the security solutions wait for the vulnerabilities in the IT infrastructures and then take action on them, depending on its nature. The approach becomes different from AI and ML-enabled tools. The AI-based systems are proactive in detecting the vulnerabilities. They can analyse the pattern and discover the loose ends that can be the potential vulnerability. By recognizing the attackers' pattern, infiltrating methods can be discovered, and it becomes easy to distinguish when and how any vulnerability would make its way to the network or system.

4.2. IMPROVING THE AUTHENTICATION

Most organizations and individuals are still dependent on the traditional method of entering the login id and password for authentication purposes. Let us face it, there are very few people who are serious about creating a unique and strong password. Over that, most people use the same passwords for all or most of their accounts. Such practices can lead organizations or individuals to serious security risks. However, with modern biometric authentication methods such as face recognition and iris recognition, login authentication has become highly secure and comfortable. The use of AI in biometrics has ensured that cybercriminals cannot hack them.

4.3. BEHAVIORIAL ANALYSIS

Another great advantage of Artificial Intelligence in cyber security is its behavioural analysis ability. AI can develop a pattern by accessing users' working methods. If any malware is introduced in the system, the working way would be changed, and that's where the AI would detect the abnormality and report it to the authorities. The abnormalities could be anything like the unusual use of the internet, change in the typing speed, increment in the background activities, and more.

4.1.1. CONTROLLING PHISHING

Phishing is one of the most common cyber-attacks used by hackers to capture the login credentials or introduce the malware into the system. Artificial Intelligence can be of great help in detecting and preventing phishing attacks.AI can detect the most common phishing sources and report them to the system to prepare for the defence against it. AI can easily recognize the difference between a fake and a legitimate website in no time. It can also analyse the phishing pattern according to the specific geographical location.

4.1.2. THREAT HUNTING

As already mentioned, the traditional security programs use signature indicators to detect threats. This technique is only effective with the already encountered attacks and becomes useless when reporting the threats that have never appeared. Using the AI, the new threats can be recognized quickly. However, with it, the false-positive cases would also increase. To eliminate the number of false-positives, both the traditional detecting method and the AI behavioural analysis detection must be combinedly used.

V. BENEFITS OF AI APPLICATION IN CYBER SECURITY

5.1. PROCESSING OF MASSIVE VOLUME OF DATA

A significant advantage of AI security is its ability to process bulky volumes of data. It does so by automating the creation of algorithms to detect security threats. The data processed covers a wide range of IT network elements. These could include files shared, emails, or visited sites. Third-party software and patterns of hacker activities are likewise included.

5.2. KEEPING YOUR CYBER SECURITY ERROR FREE

Curbing network's security errors is another benefit of artificial intelligence in cybersecurity. AI isn't tired or distracted when performing the same set of tasks. Therefore, error rates are significantly reduced compared to human efforts.

5.3. ENHANCING CYBER THREAT DETECTION AND AUTOMATION

It increases the benefits of AI in cybersecurity. A high level of detection means that the monotony of human detection of threats reduces. Thus, when cyber threats detection is automated, you swiftly find links between potential risks and act fast.

5.4. ACCELARATING DETECTION AND RESPONSE TIME

AI security is beneficial in speeding up the detection of genuine issues. You can use AI to cross-reference multiple alerts and source network data rapidly.

5.5. TACKLING ADVANCE HACKING TECHNIQUES

There are complicated and advanced hacking techniques hackers use to breach data and networks. For example, obfuscation, polymorphism, etc. These techniques are malicious and tough to identify. And coupled with the shortage of cybersecurity experts, these threats can be troubling.

5.6. SECURING AUTHENTICATION

There are business sites that require visitors to log in, fill forms, or make online payments. These businesses need an extra layer of security for the website's backend. AI security can make the authentication process more secure by physical recognition.

VI. CHALLENGES OF AI APPLICATION IN CYBER SECURITY

The active use of Artificial Intelligence and machine learning is not the only challenge that the organization and cybersecurity professionals need to face. There are others, caused by shortcomings in the current approach to security.

6.1 DISTANT INFRASTRUCTURE

Today, systems communicate across continents, sending sensitive data al over the world. These transfers don't undergo sufficient protection and are easier to break into.

6.2 MANUAL DETECTION

Human teams don't have 24/7 focus on security threats and suspicious patterns. Most of the time, systems go unmonitored.

6.3 REACTIVE OF SECURITY TEAMS

Most security experts focus on facing threats rather than predicting them.

6.4 DYNAMIC THREAT

Hackers have many strategies for hiding their locations, IPs, identities, and methods. The cybersecurity field, on the other hand, is a lot more transparent and open for research – data, created by businesses, is easily accessible by criminals.

VII. AI APPLICATION IN CYBERSECURITY: REAL LIFE EXAMPLES

7.1 SECURITY SCREENING

Security screening done by immigration officers and customs can detect people that are lying about their intentions. However, the screening process is prone to mistakes. In addition, human-based screening can lead to errors because people get tired and can be distracted easily. The United States Department of Homeland Security has developed a system called AVATAR that screens body gestures and facial expressions of people. AVATAR leverages AI and Big Data to pick up small variations of facial expressions

and body gestures that may raise suspicion. The system has a screen with a virtual face that asks questions. It monitors changes in their answers as well as differences in their voice tone. The collected data is compared against elements that indicate that someone might be lying. Passengers are flagged for further inspection if they are considered suspicious.

7.2 SECURITY AND CRIME PREVENTION

The Computer Statistics (CompStat) AI system has been in use by the police department of New York since 1995. CompStat is an early form of AI that includes organizational management, and philosophy, but depends on different software tools. The system was the first tool used for "predictive policing" and many police stations across the U.S have been using CompStat to investigate crimes since then. AI-based crime analysis tools like the California-based Armor way are using AI and game theory to predict terrorist threats. The Coast Guard also uses Armor way for port security in Los Angeles, Boston and New York.

7.3 ANALYZE THE MOBILE ENDPOINT

Google is using AI to analyse mobile endpoint threats. Organizations can use this analysis to protect the growing number of personal mobile devices. Zimperium and MobileIron announced a collaboration to help organizations adopt mobile anti-malware solutions incorporating artificial intelligence. The integration of Zimperium's AI-based threat detection with the MobileIron's compliance and security engine can address challenges like network, device, and application threats. Other vendors that offer mobile security solutions include Skycure, Lookout, and Wandera. Each vendor uses its own AI algorithm to detect potential threats.

7.4 AI-POWERED THREAT DETECTION

Commodities trader ED&F Man Holdings experienced a security incident several years ago. An independent assessment indicated that the company needed to improve its cybersecurity processes and tools. The

company looked to Cognito, Vectra's AI-based threat detection and response platform. Cognito collects and stores network metadata and enriches it with unique security insights. It uses this metadata along with machine learning techniques to detect and prioritize attacks in real time. Cognito helped ED&F Man Holdings to detect and block multiple man-in-the middle attacks, and halt a cryptomining scheme in Asia. Moreover, Cognito found command-and-control malware that had been hiding for several years.

7.5 DETECTION OF SOPHESTICATED CYBER-ATTACKS

Energy Saving Trust is an organization that is striving to reduce carbon emissions in the U.K. by 80 percent by 2050. The company was looking for an innovative cyber security technology to strengthen its overall cyber Défense strategy. This includes defending the company's critical assets, including intellectual property and sensitive client data from sophisticated cyber-attacks. After careful evaluation, the company decided to focus on Darktrace's Enterprise Immune System. Darktrace's platform is based on machine learning technology. The platform models the behaviours of every device, user, and network to learn specific patterns. Darktrace automatically identifies any anomalous behaviour and alerts the company in real time. Energy Saving Trust was able to detect numerous anomalous activities as soon as they occurred and alert the security team to carry out further investigations, while mitigating any risk posed before real damage is done

7.6 REDUCING THREAT RESPONSE TIME

A global bank faced sophisticated cyber threats and advanced attacks. The bank needed to improve its threat detection and response. The existing solution could not effectively detect and mitigate new generations of threats. The bank's security team deployed Paladon's AI-based Managed Detection and Response Service (MDR) service. Paladon's threat hunting service is based on data science and machine

learning capabilities. The bank's threat detection and response capabilities for advanced attacks were enhanced. This includes data exfiltration, advanced targeted attacks, ransomware, malware, zero-day attacks, social engineering, and encrypted attacks.

VIII.   CONCLUSION

Keeping your work data and networks secure isn't an essay task. But thankfully, you can get artificial intelligence security to fortify your cybersecurity. The benefits that come with using AI in cybersecurity are incomparable to none. From managing the bulk of threat data to detecting, responding with speed, AI does it all. Most interestingly, as AI technologies improve, so will its added benefits. Artificial Intelligence techniques are more flexible and robust than contemporary cyber security solutions. Therefore, increasing security implementation and better defend system from a growing number of advanced and complex cyber threats. we should know its limits. An Artificial Intelligence technique needs continuous human communication and training. The benefit ranges from speed and accuracy in handling large volumes of data, which is humanly impossible to handle; overall reduction in the cost of securing organizations 'valuable data and resources; and increased ROI on AI powered cyber security tools amongst others. The challenges of AI applications for cyber security include the risk of adversarial AI attacks and complacency of the human factor.

IX. FUTURE ENHANCEMENT

The findings on the degree of vulnerability of AI elements to adversarial actions have raised a lot of concerns about the security of data processing environments. AI elements disregard the conventional software analysis and introduced novel attack vectors in AI algorithms operational environments. As a result of hidden dependencies, many applications may be affected. A lot of researches are required to develop theories, engineering principles and practices when employing Artificial Intelligence as system element. We can use AI in various ways for the benefit of cyber security. In future we may have most intelligent systems than these   techniques. Even the attackers or intruders will also use the AI for attacks. Clearly, the emerging advances in data understanding, handling and illustration what is more in machine learning will greatly enhance the cyber security capability of systems that would use them.

X.   REFERENCES

[1]. Anderson, Frivoled, Valdes, "Next- Generation Intrusion Detection Expert System (NIDES)"[assessed June 30 2000]

[2]. Rosenblatt. "The Perceptron- a perceiving and recognising automaton. Report 85-460-1, Cornell natural philosophy Laboratory, 1957.

[3]. "Logic Programming for Engineering", Bratko.I, AddisonWesley, 2001.

[4]. S. A Panimalar, U.G. Pai and K.S. Khan, ―AI Techniques for Cyber Security‖, International Research Journal of Engineering and Technology, vol. 5, 3, pp. 122-124, Mar. 2018. Available: https://www.irjet.net [assessed May. 29, 2020]

[5]. T.S. Tuang. Diep.Q. B, and Zelinka. I, Artificial Intelligence in the Cyber Domain: Offense and Defense: Symmetry, 2020, 12,410. available: www.mdp.com/journal/symmetry on [assessed Apr. 20, 2020.

[6]. B. Mayo, E. Tyugu, J. Penjam. Constraint Programming. Alignment ASI Series, v. 131, Springer-Verlag. 1994.

[7]. E. Tyugu. Algorithms and Architectures of Artificial Intelligence.IOS Press. 2007.

[8]. NabaSuroor and Syed Imtiyaz Hassan, "Identifying the factors of modern day stress using machine learning".

[9]. Barika.F, K. Hadjar, and N. El-Kadhi, "ANN for mobile IDS solution," in Security and Management. Computer Applications & Robotics, vol 4, 1, pp.1-5, May 2016.

[10]. TF. Lunt, R. Jagannathan. A Prototype Real-Time Intrusion-Detection Expert System.Proc.

[11]. B. Iftikhar, A. S. Alghamdi, "Application of artificial neural network within the detection of dos attacks", 2009.

[12]. P. Norvig, S. Russell. "Artificial Intelligence: fashionable Approach", 2000

[13]. P. Pranav, ―Artificial Intelligence in cyber security,

[14]. B. Christain, D.A. Elizondo and T. Watson, ―Application of artificial neural networks and related techniques to intrusion detection‖, World Congress on Computation Intelligence, pp 949-954, 2010

[15]. E. Tyugu, ―Artificial Intelligence in Cyber Defense‖, International conference on Cyber Conflict, vol. 3, pp. 95-105, Tallinn, Estonia, Jan. 2011

[16]. W. Nadine and K. Hadas, ―Artificial Intelligence in Cybersecurity‖, Cyber, Intelligence, and Security, vol. 1, 1, pp. 103-119, Jan. 2017

[17]. S. Dima, M. Robert, B. Zvi, S. Shahar and E. Yuval, ―Using Artificial Neural Network to Detect Unknown Computer Worms‖, Neural Computing and Applications, vol.18, 7, pp. 663-674, Oct. 2009

[18]. E. H. Geoffrey, O. Simon and T. Yee-Whye, ―A Fast Learning Algorithm for Deep Belief Nets‖, Neural Computation, vol. 18, no. 7, pp. 1527-1554, 2006 .

[19]. V. Thomson, ―Cyber Attacks Could be Predicted with Artificial Intelligence‖, iTechPost,www.itechpost.com/articles/1734 7/cyberattacks-predicted-artificial-intelligen ce-help.htm, Apr. 21, 2016 [Jun. 2, 2020]

[20]. S. Franklin and A. Graesser, ―Is It an Agent, or Just a Program? A Taxonomy for Autonomous Agents‖, Third International Workshop on Agent Theories, Architectures, and Languages, no. 3, pp. 21-35, 1997

Design and Implementation of E-commerce and Supply Chain Management in Agriculture

Samhitha Kallimakula[1], Chempavathy.B[2]

[1]Department of Computer Science & Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Assistant Professor, Computer Science & Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This paper provides a strategy of best use of technological know-how and innovation to the back bone of our country, farmers. Other streams have utilized technological know-how and benefitted to an increased extent and running their business. Applying the supply chain administration of agricultural products and along with e-commerce will be avoiding the intermediate layers and really helpful to producers additionally preserve transparency in the process.

Keywords - e-commerce and agriculture supply chain management

I. INTRODUCTION

An agribusiness item store network is the gathering of associations, individuals, innovation, exercises, data and sources worried in moving item from supplier to client. According to the definition, it is obvious that all the data is accessible to any important partner. This paper, focuses on how well the farmers can benefitted and consumers can be satisfied by leveraging the e-commerce application following the supply chain principles applied to the agriculture and at the same time making best use of technology to provide transparency to all stakeholders involved.

II. THE PROPOSED MODEL OF EBIZ SUPPLY CHAIN MANAGEMENT

The store network combination model of horticulture item coordination's in the E-trade is to naturally turn away the transitional layers like retailers, vendors or representatives of farming item through the E-business data foundation of inventory network dependent on the data organization and the makers, providers and customers of the agribusiness item structure the incorporated activity of creation, supply and offer of farming item through the data stage and the consistent association through ebiz framework. The agribusiness item store network model under the E-trade climate is appeared as in Fig. I.
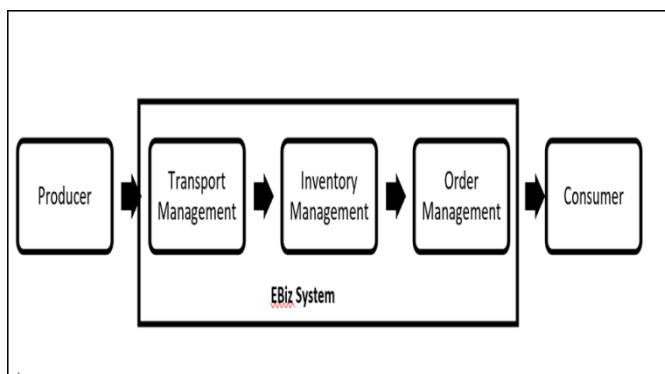
Fig. 1. supply chain model for agriculture product e-commerce system

The creation administrative association, deals association and market administrative association of farming item likewise can straightforwardly administer the result of makers, market access and quality security of the agribusiness item through the data stage and issue the most recent worldwide and home-grown norms of horticulture item through the data stage to manage the creation simultaneously and the purchasers can enquire the quality wellbeing of the purchased agribusiness item through the organization terminal of the data stage and furthermore can follow back the delivering region, which basically ensures the rights and interests of the customers and furthermore is ideal for the brand creation and insurance of farming item. It can make the entire stream interaction of farming item high-compelling, coordinative and systematic to finish the coordination's circulation capacity of the agribusiness item inventory network by the outsider coordination's dissemination focus subsequently to diminish the misfortune and save the expense. Warning from ebiz, issue the creation data to the ranchers as per the interest data according to our market interest and the deal terminal on the E-business data stage and the creation and handling endeavours just as the middle person associations buy the horticulture item through the orders. Thusly, the creation visual impairment of the ranchers is diminished and the assurance is accommodated the selling lines of the horticulture item simultaneously, which not just brings down the coordination's cost of

farming item and decreases the deficiency of agribusiness item yet in addition is ideal for the industrialization of rural creation.

III. MODULE DESIGN OF E-BIZ SYSTEM

There are five modules in agribusiness item inventory network internet business framework. They are Portal Website, Order Management, Transportation Management, Inventory Management and Interface involves entry site and versatile application. Figure 1 shows the design of fundamental modules in horticulture item inventory network ebiz framework. This framework incorporates significant exercises like buying, stock stream booking and control, coordination's creation coordination, transportation frameworks activity and foundation, client support, request satisfaction and dissemination offices the executives. The ebiz system comprised of different modules to meet the need of agriculture product supply chain model under the E-commerce environment is shown as in Fig.2
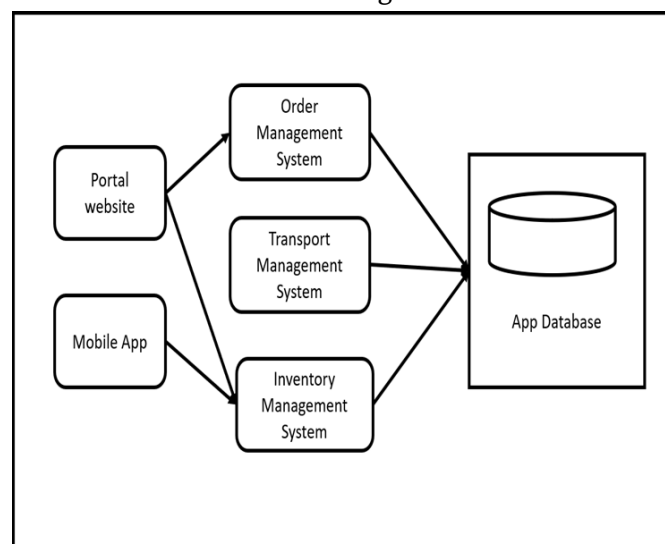
Fig. 2. Module design of agriculture product e-commerce system

A. Interface System

1. **Portal Website:** Entryway Website might be a stage to supply general data for rural production network accomplices and offers a genuine human-

PC connection interface. The clients can acquire data of environment and agribusiness news from anyplace in the web application. Current cost and brand qualities of horticulture item, can be given to the clients through entrance. Also, it serves the other purpose of placing orders from anywhere in India, provided the user is connected to internet and can able to access our portal.

2. **Mobile Application:** This interface is needed for the field operator part of Transport Management System responsible for the collection of agricultural products from farmers and move it to the cold storage units. This application comprises of sharing the current price of a selected product and accordingly the total price is to be paid by the farmer. Now, as part of collecting the products field operator weigh in the products which are collecting from the farmer and records an entry to the ebiz inventory sub system, so that further this helps in generating an invoice to the farmer once goods are confirmed by inventory management. A unique reference is generated and sent to the farmer, to keep track of their orders

B. Inventory Management Module

Stock Management Subsystem involves stock data the executives of info and yield, items and stock status data the board, items bundle and heaping expenses the executives, declaration the board and interaction observing, bookkeeping articulation the board and files upkeep. It is a key a piece of the farming production network and basically means to control the development and capacity of materials inside a distribution centre and interaction the related exchanges. The frameworks create an extraordinary standardized tag for every item and keep up the rundown in our ebiz data set to productively screen the progression of items. Whenever information has been gathered, there is consistently correspondence channel exists between request the board and stock subsystem. The information base would then be able

to give valuable reports about the situation with items inside the stockroom.

C. Order Management Module

To acknowledge constant request the board, a horticultural items notoriety ensure framework ought to be set up. This framework can be utilized to help track horticultural items, methods for rural creation providers, and shoppers, and the reason for existing is to improve the readiness of activity later on slowly. Customers can place the orders of their choice from the list of products available in our inventory displayed to the user. Once order is placed, order management system takes care of order tracking and same status showed to the customer any time. The information base would then be able to give valuable reports about the situation with orders for any day in the ebiz framework.

D. Transportation Management Module

Transportation Management Subsystem generally maintains the list of stock to be collected on a particular day and accordingly it is completed.It integrates with our inventory management source system. By contrasting the transportation cost and time necessity of the items, the target of the framework is to improve the course of the item transportation in order to control and reduce the expense of the entire agrarian store network.

IV. CONCLUSION

Today, the advancement of modem inventory network the board of farming items is as yet in developing stage in India and degree for improvement exists using web based business. To take care of these issues, our inventory network ebiz framework helps in making the store network the board, advance the improvement of agribusiness item data in India and structure new systems and new proportions of farming item flow.

V. REFERENCES

[1]. Applications of Supply Chain Management and E-Commerce Research; Geunes, J., Akçali, E., Pardalos, P., Romeijn, H.E., Shen, Z.-J.M. (Eds.)

[2]. The Development and Strategies of Agricultural Supply Chain under the Background of Economic Globalization", Northern Economy and Trade, vol.12, Nov.2007, ppI5-17.

[3]. Wang Ning and Huang Liping,"A Study on Agricultural Product Logistics Supply Chain Management Mode Based on Information Network", Agricultural Modernization, vol.2, Feb.2007, pp26

[4]. Ye Weiyuan,"On the Construction of Agriculture Product Logistics System Based on the Third-party Logistics" Modem Logistics. vol. IO,Ju I .2008,pp44-46

Smart Garbage Monitoring System Using IoT

Nehaa R[1], R Kiruthiga[1], Saara P[1], Vaishali M Deshmukh[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Rd, near Marathalli, Kaverappa Layout, Kadubeesanahalli, Kadabeesanahalli, Bengaluru, Karnataka 560103, India

ABSTRACT

Many times, in our city we see that the garbage bins or dustbins placed at public places are overloaded. It creates unhygienic conditions for people as well as ugliness to that place leaving bad smell. To avoid all such situations, we are going to implement a project called IoT Based Smart Garbage and Waste Collection bins. These dustbins are interfaced with micro controller-based system having Ultrasonic wireless systems along with central system showing level and current status of garbage, on web browser with html page by Wi-Fi. Hence the status will be updated on to the web page. Major part of our project depends upon the working of the Wi-Fi module; essential for its implementation. The main aim of this project is to reduce human resources and efforts along with the enhancement of a smart city vision.

Keywords — Garbage monitoring, Internet of things, Arduino, Ultrasonic sensors, Power supply, Collecting data, Store data, Webpage, Wi-Fi module

I. INTRODUCTION

Internet and its applications have become an integral part of today's human lifestyle. It has become an essential tool in every aspect. Due to the tremendous demand and necessity, researchers went beyond connecting just computers into the web. These researches led to the birth of a sensational gizmo, Internet of Things (IoT). Communication over the internet has grown from user - user interaction to device – device interactions these days. The IoT concepts were proposed years back but still it's in the initial stage of commercial deployment. Home automation industry and transportation industries are seeing rapid growth with IoT. Yet not many articles have been published in this field of study. This paper aims in structuring a state-of-the-art review on IoT.

The technology, history and applications have been discussed briefly along with various statistics.

Since most of the process is done through the internet, we must have an active high speed internet connection. The technology can be simply explained as a connection between humans-computers-things. All the equipment's we use in our day-to-day life can be controlled and monitored using the IoT. A majority of process is done with the help of sensors in IoT. Sensors are deployed everywhere and these sensors convert raw physical data into digital signals and transmits them to its control center. By this way we can monitor environment changes remotely from any part of the world via internet. This systems architecture would be based on context of operations and processes in real-time scenarios.

Smart collection bin works in the similar manner with the combination of sensors namely weight

sensor and IR sensor that indicates its weight and different levels respectively. The IR sensors will show us the various levels of garbage in the dustbins and also the weight sensor gets activated to send its output ahead when its threshold level is crossed. These details are further given of the microcontroller (ARM LPC2148) and the controller gives the details to the transmitter module (Wi- Fi module). At the receiver section a mobile handset is needed to be connected to the Wi-Fi router so the details of the garbage bin is displayed onto the HTML page in web browser of our mobile handset.

The noted advantages of using Garbage monitoring system are as follows:

i. It saves time and money by using smart waste collection bins and systems equipped with fill level sensors. As smart transport vehicles go only to the filled containers or bins. It reduces infrastructure, operating and maintenance costs by up to 30%.

ii. It decreases traffic flow and consecutively noise due to less air pollution as result of less waste collection vehicles on the roads. This has become possible due to two-way communication between smart dustbins and service operators.

iii. It keeps our surroundings clean and green and free from bad odour of wastes, emphasizes on healthy environment and keep cities more beautiful.

iv. Applying smart waste management process to the city optimizes management, resources and costs which makes it a "smart city".

v. It helps administration to generate extra revenue by advertisements on smart devices.

II. RELATED WORK

Literature survey is the most important step in software development process currently. Before developing the tool, it is very necessary and important to determine the time factor, economy and company strength. Once these things are satisfied and requirements are met, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support and accessories. This support can be obtained from senior programmers, from books or from websites as references. Before building the system all the above points are taken into consideration to the account for developing the proposed system. Hence, it is very important to have a proper literature survey done before going ahead with any of the project work. Below lists some of the surveys that have been performed in order to obtain the results.
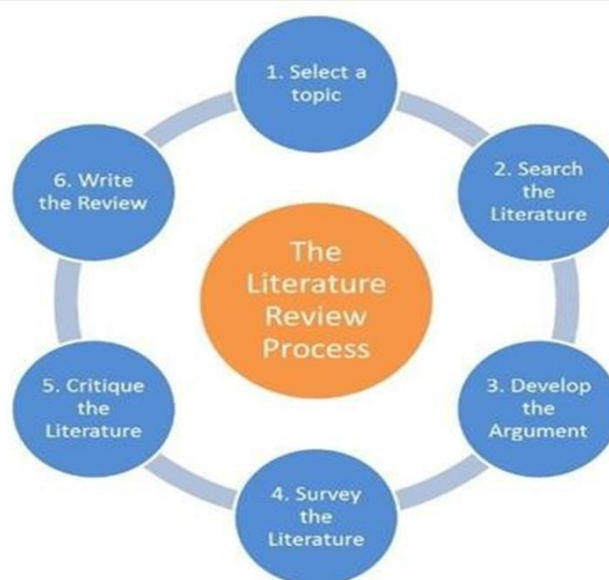
Fig 1: Process of Literature Review

The existing system in current days is very time consuming and less effective, trucks go and empty containers whether they are full or not which leads to lots of fuel waste. High costs. Unhygienic Environment and look of the city. Bad smell spreads around the surroundings and may cause illness to human beings and animals.

A state-of-the-Art review on Internet of Things. Gives the idea of IoT subject as well as the related information. Training should be provided to people.[1] Internet of Things: Challenges and state-of-the-art

solutions in the internet scale sensor management and mobile analytics. Gives details about mobile analysis and sensor information management that helps in data segregation of various dustbins. Low data speed and shorter range.[2] Top-k Query based dynamic scheduling for IoT- enabled small city waste collection. Gives us the conceptual understanding of dynamic scheduling required for cleaning the dustbins and top-k leads us to priority-based cleaning. High initial cost.[3]

Smart garbage System Management, provides additional designs and working, high initial cost.[5] IoT based Smart Garbage system for efficient food waste management, overview of smart garbage monitoring system. Requires more waste bins.[6] Smart Bin, Applications and uses of smart garbage monitoring system in maintaining a friendly environment. Training has to be provided. Reduces man power. [11]

III. METHODOLOGY

An ultrasonic sensor will be placed on the interior side on the trash bin, the one side of the sensor facing the solid trash. As the amount or level of trash increases, the distance between the sensor and the trash in the bin decreases. This current level of data will be sent to our microcontroller. Microcontroller then processes the data sent by the sensor and sends to HTML page connected to the web browser. At the receiver end any device will be connected which displays the amount of thrash or the current level of the trash in the bin.
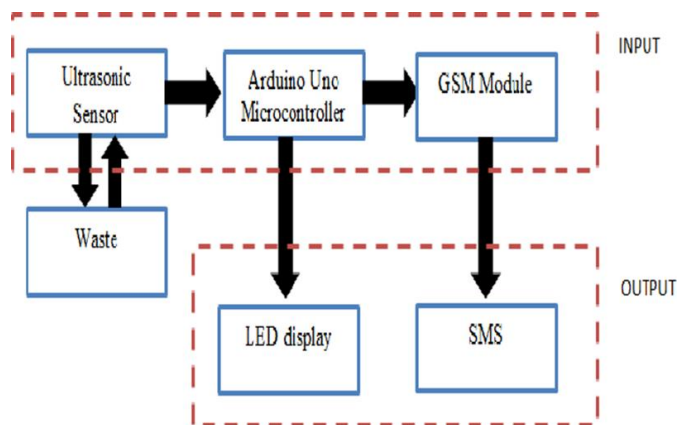
Fig 2: Representation of garbage monitoring

IV. RESULT AND DISCUSSION

Ultrasonic transducers/sensors are transducers that convert ultrasound waves to electrical signals or vice versa.

The sensors are placed on the top of the bin facing the trash in the bin they measure the level of trash based on the distance between the trash in the bin and the sensor placed at the top.

If the trash thrown in the bin is very less or little there will be huge distance between the sensor and the trash as the amount of trash is less and is measured in cm (centimeter), here in the below example the garbage level is 24cm. The value is displayed on the Arduino serial monitor window and then the data is transmitted to the web page via Wi-Fi module.

Fig 3: Example 1 The arduio serial monitor showing the garbage level is at 24cm.

If the trash thrown in the bin is half or little less than half, then the distance between the sensor and the trash will be reduced and is measured in cm (centimeter), here in the below example the garbage level is 12cm. The value is displayed on the Arduino serial monitor window and then the data is transmitted to the web page via Wi-Fi module.

Fig 4: Example 2 The arduio serial monitor showing the garbage level is at 12cm.

If the trash thrown in the bin more than half or almost full, then the distance between the sensor and the trash will be reduced to a great extent as the bin is almost filled and is measured in cm (centimeter), here in the below example the garbage level is 10cm. The value is displayed on the Arduino serial monitor window and then the data is transmitted to the web page via Wi-Fi module.

Fig 3: Example 3 The arduio serial monitor showing the garbage level is at 10cm..

V. CONCLUSIONS

Smart Garbage Management is very much important towards having the city clean and hygienic. Traditional Garbage management employing human is not very effective and requires more man power with no proper supervision. Research has been done in employing the IoT based technology in monitoring the status of the garbage bin towards collecting the trash once the threshold level is reached. So IoT based Smart Garbage Management System has been developed as a prototype where ultrasonic sensors are fitted in the garbage bin for monitoring the depth of bin and accordingly once the threshold level is reached alarm is buzzed. This information is updated in HTML page viewed through the web browser. This web page also sends all information and updates to garbage collection vehicles. The web page is built to show the present status to the user monitoring it. The web page gives a view of the garbage bin and it also highlights the color in order to show the level of garbage collected in the bins. The system puts on the buzzer when the level of garbage in the bin crosses the limit set in the system. Thus this system helps to keep the city clean by informing about the present garbage levels in the bins by providing the graphical image of the bins via a web page.

VI. REFERENCES

[1]. P.Suresh1J. Vijay Daniel2, Dr.V.Parthasarathy4" A state of the art review on the Internet of Things (IoT)" International Conference on Science, Engineering and Management Research (ICSEMR 2014)

[2]. Arkady Zaslavsky, Dimitrios Georgakopoulos" Internet of Things: Challenges and State-of-the-art solutions in Internet- scale Sensor Information Management and Mobile Analytics" 2015 16th IEEE International Conference on Mobile Data Management

[3]. Theodoros.Anagnostopoulos1,Arkady.Zaslavsky2 ,1, Alexey Medvedev1, Sergei Khoruzhnicov1" Top–k Query based Dynamic Scheduling for IoT-enabled Smart City Waste Collection" 2015 16th IEEE International Conference on Mobile Data Management.

[4]. "City Garbage collection indicator using RF (Zigbee) and GSM technology"

[5]. Vikrant Bhor, Pankaj Morajkar, Maheshwar Gurav, Dishant Pandya4 "Smart Garbage Management System" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS031175 Vol. 4 Issue 03, March-2015

[6]. Insung Hong, Sunghoi Park, Beomseok Lee, Jaekeun Lee, Daebeom Jeong, and Sehyun Park, "IoT-Based Smart Garbage System for Efficient Food Waste Management", The Scientific World Journal Volume 2014 (2014), Article ID 646953

[7]. Marian Look, "Trash Plant: India", earth911B.

[8]. Basic Feature, "Solid waste Management Project by MCGM

[9]. Pavithra "Smart Trash system: An Application using ZigBee" International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 8, October 2015

[10]. "Smart Garbage Monitoring System" By Arul Anitha, Dr. L. Arockiam.

[11]. "Smart Bin" By Musafa M.R, K.V Azir.

[12]. "IoT based Smart Garbage System for efficient food Waste Management by Insung Hong, Sunghoi Park, Jaekeun Lee.

Carbon-Neutral College Campus

Akilesh Murugan, Sounak Koner, Gobind Kumar Thakur, Ms. Shanmuga Priya

Department of Computer Science and Engineering, Bangalore, Karnataka, India

ABSTRACT

In this paper, we present a machine learning approach to provide a probable solution to reduce global warming. We provide our system with verified databases having information of carbon emission levels along with time stamps. Although our learning approach uses only two features, we are able to achieve quite competitive accuracy. The result indicates different action points we can use to reduce carbon emission and over a period of time, achieve carbon emission neutrality at any organization.

Keywords — machine learning, carbon footprint, carbon emission, linear regression, time series, ARIMA model

I. INTRODUCTION

Mankind is facing the biggest threat of existence, caused by our actions. The consequences of this are global air pollution and global warming. Carbon dioxide levels can increase surface temperature by 1.5-3°C in next half decade [1]. Even if eradicating the use of fossil fuel is not viable, with the use of machine learning and artificial intelligence a systematic software/ application can be developed, enabling human race to reduce carbon emissions and moving towards green energy. However, this starts with ground work of identifying various factors responsible for carbon emission and studying them. About the levels of emission and how far it has been scaled. This research makes up the raw data which is used later in the models for prediction.

Global warming has gained a lot of attention on a nation- wide growth. Added as sustainable development goal (SDG) with various limitations on businesses and transport in-order to take control of carbon emission rates. The SDG is focused on development and achieving targets. Central to climate actions is another agreement which is United Nations Framework Convention on Climate Change (UNFCCC) [2].

Another strand of work is collecting data from resources and people on a regular time-gap.

In this paper, we address the development of software that takes in data through multiple databases and providing a set of action events that needs to be followed to cut down the level of carbon emissions. As this works on multiple machine learning algorithms, we have a mathematical backing about why various steps are followed. If the same idea and working reaches to all schools, colleges and business organization, we can cut down substantial carbon emission levels.

As students of final year of graduation, we have taken college campus as the focus, in addition to educational institutions not-being-a-focus-area of sustainability in India. There are references available for colleges / universities outside India achieving carbon-neutrality, the college team took this as an attempt and try to determine a-way-forward with kick-starting developing a software program and find relevant and

possible management interventions to help the educational institutions to reach carbon neutrality.

II. METHOD

2.1 Data collection

Our solution providing model has datasets on its base with which predictive models work.

The first verified dataset is downloaded from online resource having two features (year, carbon emission value). These values correspond the total carbon emission level each year. Having 160 entries since the year 1860, we started working on the basic model structure. We also started collecting, second dataset, data entries from college students in time-gaps of few days to identify various factors which form a major component of the model.

Now having a continuous data flow given by students, we focused on electricity consumption by the college. Consuming electric power indirectly gives rise to carbon emission (burning of fossil fuels). Thus, identifying components that use electric power as source.

This included light bulbs, computer systems, lab equipment to hostel electric use. After this we found out which components were used for how long in a day and how many days in a week, etc. So, we extrapolated this data and came up with a carbon emission value that is added to the entries in the second dataset.

2.2 Machine Learning Approach

Our system uses two predictive models to study the data and give us a possible set of action points to be followed. They include linear regression and time series analysis. Linear regression is a linear approach to model relationship between dependent and independent variables. The base is $Y = mx + c$, linear equation. Line representing this equation is placed on the graph. This model is only used for the first dataset as the growth of carbon emission since 1960 is almost linear. We achieved the coefficient of determination

at 0.925, giving an acceptable accuracy prediction with respective to the model.

The second predictive model uses time analysis. Giving continuous data line growth. Time series works on data points collected at constant time intervals. Unlike linear regression observations being independent doesn't hold true. The other specialty of time series is identifying seasonal trends that helps in identifying specific time frames that have similar behaviors. Hence, we use this model with the second dataset. Data is continuously analyzed for its growth, direction and repetition which provides much accurate predictions on live data feed.

Before analysis, data is cleaned. Meaning, there are no missing data points, as this will give a misleading analysis. Then outliers are removed to obtain much accurate analysis. Once data processing is done, datasets are divided into training and testing sets. As the names suggest, training datasets train the model while test datasets are used to give predictive analysis. Python programming language is used for software execution model while Kivy is used for UI development.

2.3 Providing Solution

After data organization was done, we started working on the module that provides solution. We take 0.002% of the predicted value from linear regression (dataset 1) and 80% of the predicted value from ARIMA model (time series, dataset 2). The action events that are used in providing solutions are:

- Planting
- Travelling in eco-friendly manner
- Investing on green energy source
- Changing hardware / technology to products or processes that consume lower energy

These activities can be given to individuals for implementation to achieve the desired goals of reduced carbon emission.

Formula to be used: Total value of carbon emissions
                     Number of action events

Ex: Say we have to reduce 400 units of carbon emission over a period of 2 years.

The number of action events available are 4. By using the formula, to reduce emission/ increase equivalent green activities of 100 units using each action event in 2 years.

By knowing the average oxygen generated by a vegetation (example, planting trees) grown in 2 years, the quantity of plants to be sowed is analyzed.

By identifying the transport method of all students and faculties (from self-driven two or four wheelers to public transport of bus or train), change in transport routine will be specified.

Understanding how much emission a single unit of electric power generation we can invest in generating green energy such as solar panels and cut down fossil fuel power consumption.

Changing simple electronic hardware such as lighting (example from changing the light bulbs to LED), various lab components and optimizing the use will indirectly reduce carbon emissions.

If the availability of a particular action event is not feasible, the system can share the same with other action events.

III. RESULTS AND DISCUSSIONS

The below plotted graph, Fig.1, represents year in x-axis and value of carbon emission along y-axis. The line is placed by analyzing the data population (data density as a specified point and the direction the plotted points follow).

Plotted points from year 1860 to 1970 show a linear growth. Data points from 1990 to 2013 represents linear growth. The significant change from a plateau to an exponential rise is due to the fact of growth in industrialization became much stronger after the year 1970. 1970 marks the beginning of industrialization in India. When these states of growth are combined under linear regression, the line gives us a coefficient of determination of 0.92. This is an acceptable accuracy level, as a linear model.

This model cannot be used when there are fluctuations in data points. The model is only used while there are fewer fluctuations and a general growth of dependent variable (here, carbon emission value) is known. Since the data set is not a man-made simple and controlled process, but combination of man-made (industrialization) and natural processes, the data fluctuation is natural as the seasons and rain etc would also seriously affect carbon emission measurement levels.

As there is only one variable plotted, the equation looks like $Y = B_0 + B_1x$. If there are more variables, then the equation would be $Y = B_0 + B_1x_1 + \ldots + B_nX_n$ where X_1 to X_n are the inputs.
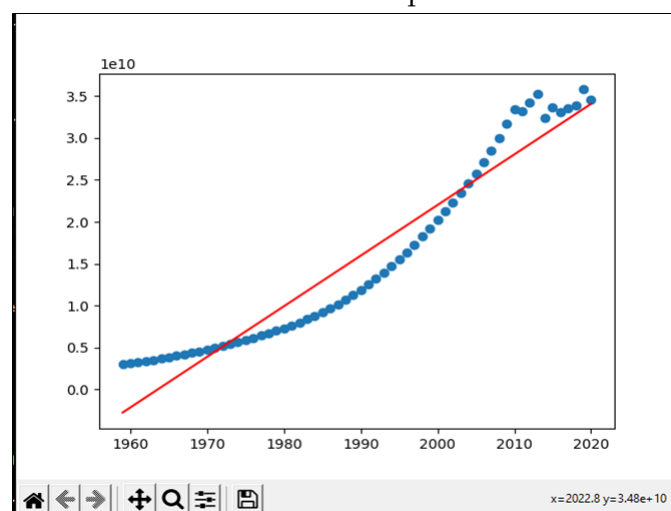
Fig. 1 Linear Regression

Fig.2 represents time analysis based on ARIMA model (Auto-Regressive Integrated Moving Average). This model is a combination of two other models, namely auto regression and moving average. This model is used in statistics and econometrics to measure events that happen over a period of time. The dataset used in below graph is a sub-section of first dataset. The year 1960 marks the beginning of significant industrialization.
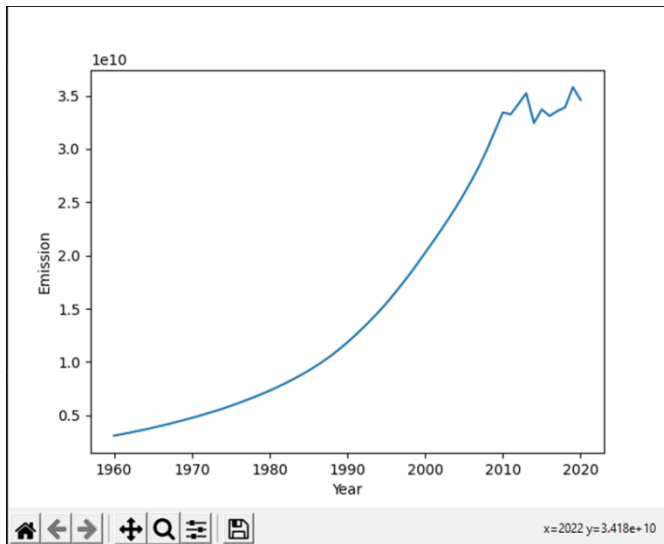
Fig. 2 Time Series

To determine if the dataset we are dealing with is stationary or not, we plotted rolling mean and standard deviation of the time series.
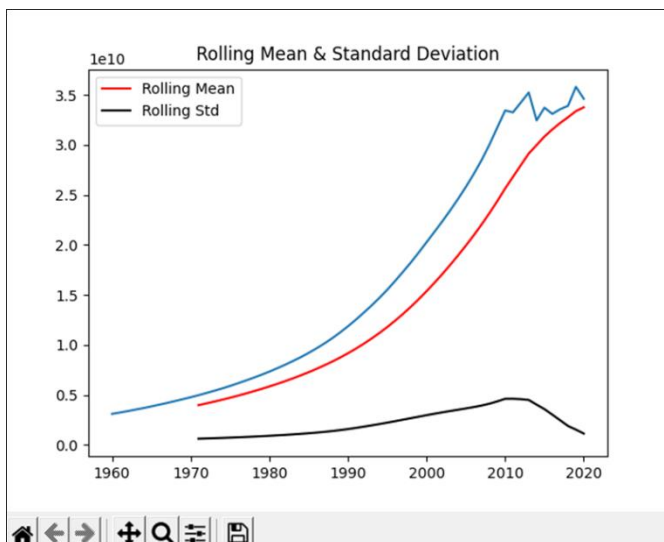
Fig. 3 Rolling mean and standard deviation

Rolling mean moves in an upward trend, indicating the dataset to be non-stationary. Further, while using Dickey-fuller test the output confirmed the dataset is non stationary as statistical value > critical value (failed to reject null hypothesis).

To convert dataset to a stationary dataset we extrapolated the dataset by applying log scale. Using this we reduced the exponential curve to an approximate straight line. Thus, getting the moving average more stationary as shown below in Fig. 4.
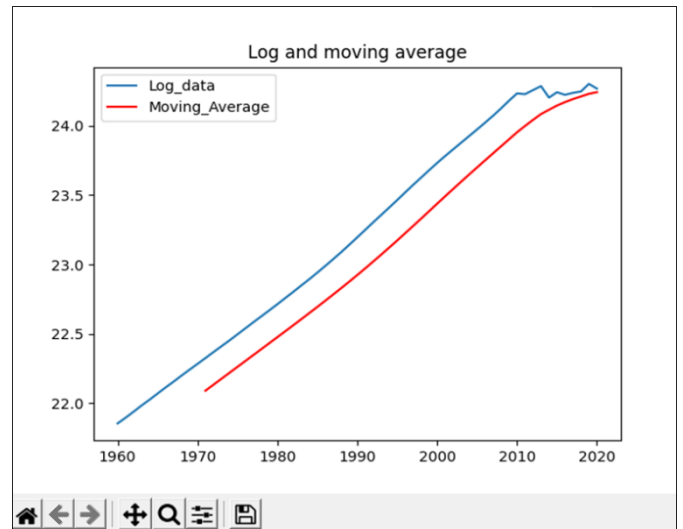
Fig. 4 Log and Moving Average

Then with the use of exponentially weighted moving average, plotting much accurate growth was found as shown in Fig.5.
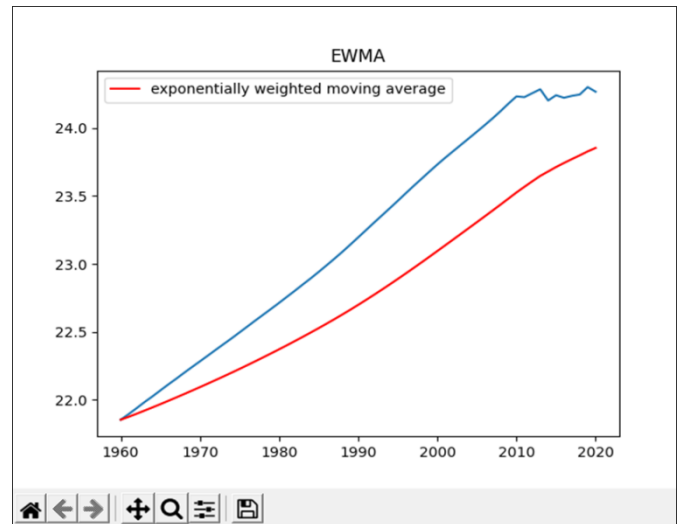
Fig. 5 Exponentially weighted moving average

By using this approach, weights for older data points are reduced and newer data points gain more weights, hence, giving more importance to recent activities. Now with data shifting we arrive at much stationary dataset. Given below (Fig.6) is the data representation that has achieved alternative hypothesis, achieving stationarity.
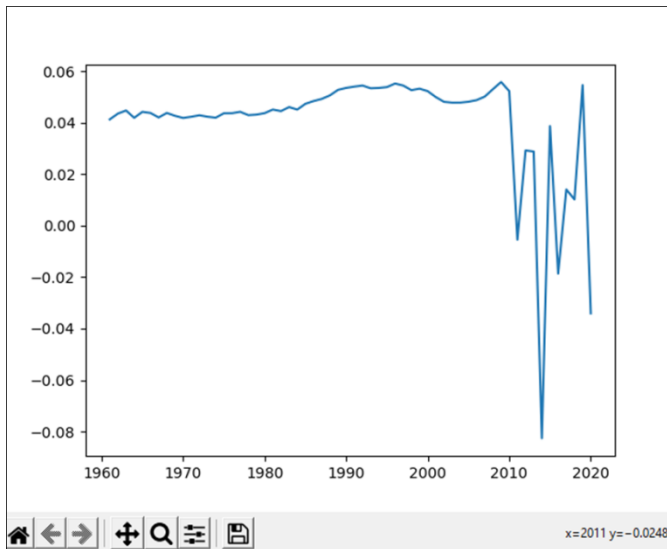
Fig. 6 Stationary data

Basic necessity for prediction is met. For time series, we find the trend, seasonality and residuals.

Trend: the direction and altitude of growth

Seasonality: pattern/ set of patterns that seem to repeat

Residuals: abnormal data points that change the course of growth

(Fig.7) represents these factors.
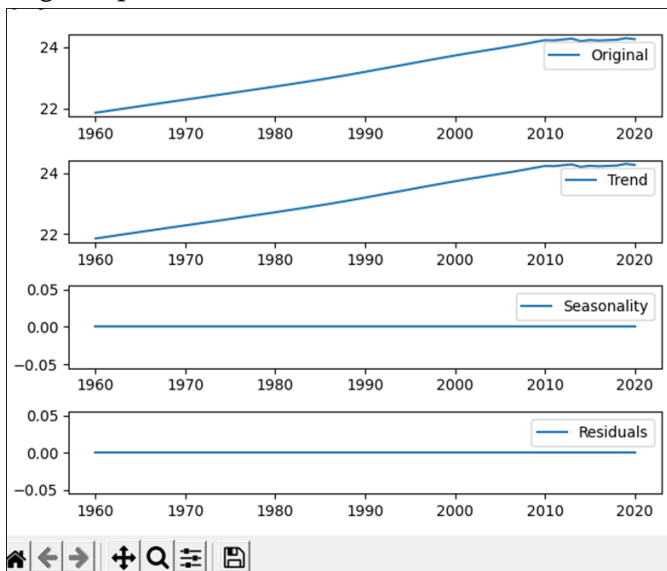
Fig. 7 Trend, Seasonality & Residuals

To find p and q values for the model we plot partial auto-correlation and auto-correlation functions respectively. p represents autoregressive terms and q is moving average terms. Fig.8 represents these functions.
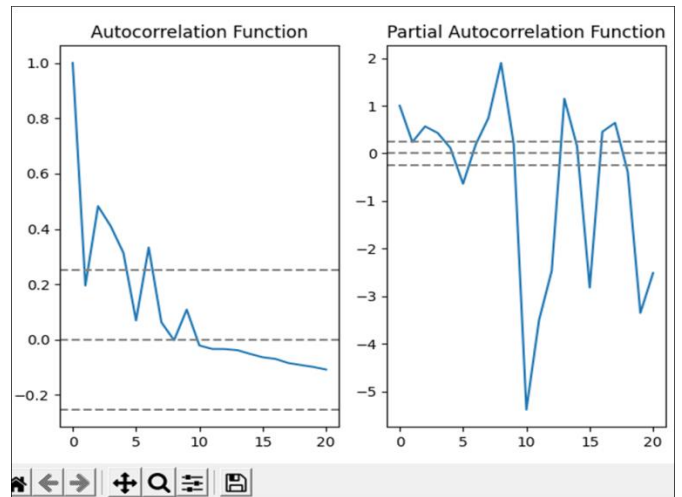
Fig. 8 Autocorrelation (q) and Partial Autocorrelation (p)

Also represented as AR(p) MA(q). Fig.9 represents AR model fitted onto the stationary data. Fig.10 represents MA model fitted onto the stationary data.
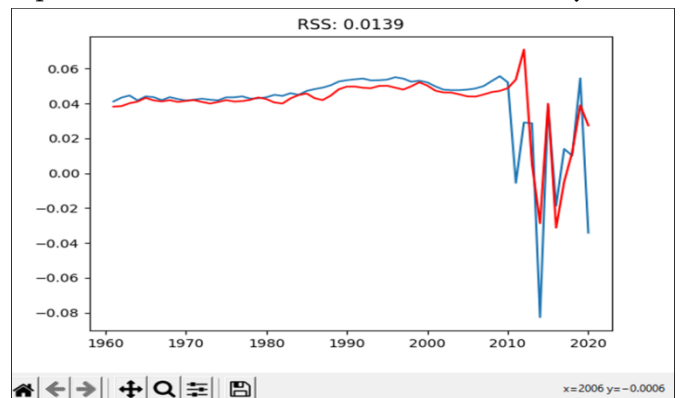
Fig. 9 AR model
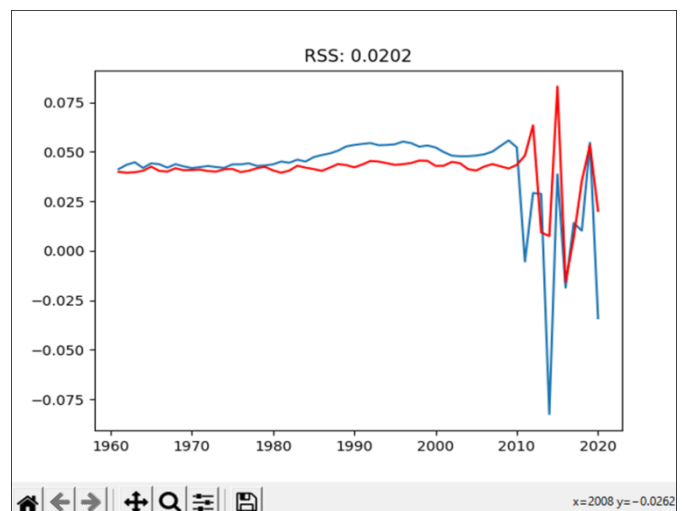
Fig. 10 MA model

RSS represents residual sum of squares. The lower the value better is the accuracy and prediction. By combining AR and MA models we predict the future carbon emission as shown below in Fig.11.
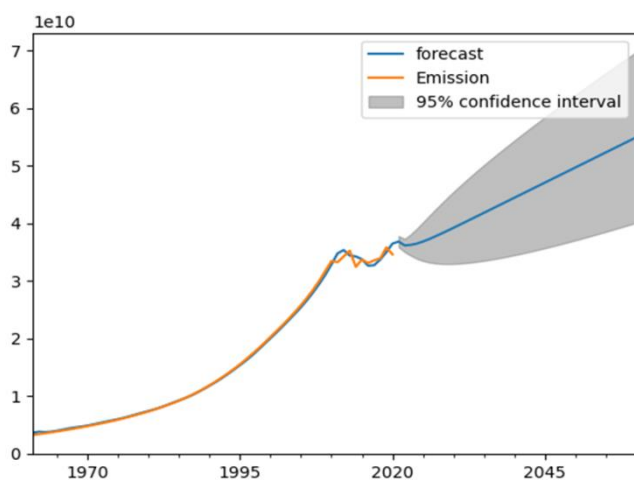
Fig. 11 Predicted data

With this system any organization/ educational institute can follow a systemic solution and achieve carbon neutrality. The rise in carbon remissions is exponential and thus big organizations working with governments and private institution to change the way businesses are functioning and their goals. Indian government has set targets for carbon emission reduction by placing a PAT scheme and promote companies to become more sustainable [3] and motivate such organizations to sell their energy saving to those who are not able to meet their energy saving targets through a trading platform [4]. This is in line with globally recognized Kyoto Protocol or Chicago Climate Exchange mechanisms [5,6]. It has become mandatory for large designated power consuming industries. This has also become a means of marketing to such organizations to claim that they are becoming more "sustainable" focused business organizations.

3.1 Future Work

3.1.1   The next work is to complete the collection of datasets to 100 entries which is already work-in-progress. This is an iterative development and will always have changes and updates, as we learn more about the changing behaviors of global warming and technology changes. But, as the next update, it would be important to access all the data from a cloud/ server as data is the most essential component of this system. We can't take the possibility of the data being tampered. The system is accessed only by admin. Hence, we need to build UI for the user to have reference on updates and plans. Once UI is available to users, data must be highly encrypted.

3.1.2   Educational institutions, particularly, can encourage students to form student clubs which can take responsibility to make awareness to stakeholders of the educational institutions to build sustainable methods to keep reducing carbon emission. le for active participation of students. Active students who perform necessary action events can be motivated through awarding them with badges/ role promotion/ work with core team. So that students are motivated and the number of participants keep growing.

3.1.3   For other business organizations, universities etc., maybe, it is essential to make amends to the program to modify input variables which will help them accurately determine carbon emissions through extrapolation to the future and also modify the action events, as human race is learning more beyond planting trees, solar / wind energy conversions and other technological changes.

IV. CONCLUSIONS

In this paper, we present a machine learning approach to reduce carbon emissions to make the college / educational institution carbon-neutral. The plotted graph predicts the direction the trend will move. The output of the program in terms action events are provided as a possible solution that can be followed to achieve carbon neutrality.

As businesses have already realized that spending / investing money on such sustainability action items

pays for itself, the educational institutions will also not be any different.

V. ACKNOWLEDGEMENT

VI. REFERENCES

[1]. https://climate.nasa.gov/news/2865/a-degree-of-concern-why-global- temperatures-matter/

[2]. https://unfccc.int/topics/action-on-climate-and-sdgs/action-on-climate- and-sdgs

[3]. https://beeindia.gov.in/content/pat-3

[4]. https://www.iexindia.com/

[5]. https://en.wikipedia.org/wiki/Chicago_Climate_Exchange

[6]. https://unfccc.int/process/the-kyoto-protocol/mechanisms/emissions- trading

Near Field Communication Based Attendance System

Jervis Dias[1], Neil Dias[1], Juviane D'costa[1], Mrinmoyee Mukherjee[1]

[1]Department of Information Technology, St. Francis Institute of Technology, Mumbai-400103, Maharashtra,

India

ABSTRACT

Registering for attendance in education environments is a highly demanding activity as a result of increasing number of students. In most colleges attendance of students plays an important role since the students are graded accordingly and taking attendance for so many students can be painstaking. Thus this project focuses on developing a smart attendance system using NFC that will simplify the attendance process, by simply touching the student's NFC enabled ID card to the lecturer's NFC based mobile device in the class. The attendance will get updated simultaneously to the database during run time. The system is based on NFC Technology and run on mobile as an application.

Keywords — Attendance, Mobile Application, Near Field Communication (NFC).

I. INTRODUCTION

The attendance process normally involves circulating a paper for the students to register their names, or the lecturer calling the names and registering the students either on a paper. It is a time-consuming process as it involves calling a particular student and then filling information accordingly. And after that the student's attendance has to be updated to the college database by the teacher manually, which is again tedious. Imagining the number of students to be from 50 and above, a great portion of the lecture time will be wasted performing this process [1]. So, NFC Based Attendance System is implemented. To solve all the issues, we have implemented an android application based on Near Field Communication (NFC) that will mark the student's attendance with the NFC tag in the ID card; thus, updating the database in real time efficiently. The benefits of this system include eliminating the chance of losing attendance data,

different attendance reports can be easily generated by a click of mouse, simplifying the decision making process related to attendance, etc. One of the major distinct characteristics of our proposed system is that the hardware required is minimal.

II. RELATED WORK

In 2008, Nucleus proposed the use of a computerized attendance system, which can eliminate human involvement, human data entry mistake, repetitive work. This system is going to increase productivity, reduced payroll error, and reduced payroll inflation, reduced overtime, retirement of legacy systems, Elimination of paper costs, and which can provide all the reports on demand. In this system, faculty has to take attendance manually, only these records have to be entered into the computerized system. But in this problem of data entry may occur [2]. In another paper the authors proposed a system comprised of using the

Viola Jones algorithm for detecting the human faces and then the detected face is re-sized to the required size, this re-sized face is further processed by using linear stretch contrast enhancement and finally it is recognized using a simple PCA / LDA. Once recognition is done, automatically attendance will be updated in an Excel Sheet along with his name, date and time. An HTML file is automatically updated by our system so that a remote authenticated user can access the attendance file. The main problem in this system is recognized face has to be compared with all the entries in the database [1].In another paper by BIS, they present a commercial system based on RFID for attendance management for schools and colleges. The system can send SMS and email alert to parents/guardians of the students automatically. The student will register at the gate by touching RFID device with their RFID tag and send the data to BISAM server in the school.The server will process the attendance data and send an SMS to the parents/guardians of the absentee student through BISAM SMS gateway server. The system also has Time Manager Software for managing employee's attendance and HR related functionalities. . The problem in this research is there is verification is not done. So proxy attendance may be marked [3]. In 2013, Vishal Bhalla et al, have proposed the attendance system which can take attendance using Bluetooth.In this project,attendance is being taken using instructor's mobile phone. Application software is installed in instructor's mobile phone enables it to query students mobile phone via Bluetooth connection and through transfer of students mobile phone Media Access Control (MAC) addresses to the instructor's mobile phone, presence of the student can be confirmed. The problem of this proposed system is student's phone is required for attendance [4]. In 2010, Seifedine Kadry and Mohamad Smaili has proposed one system. In this paper, a wireless iris recognition attendance management system is designed and implemented using Daugman's algorithm (Daugman 2003). This system based

biometrics and wireless technique solves problem of spurious attendance and the trouble of laying the corresponding network. It can make the users attendances more easily and effectively. Main problem in this system is it is too expensive and it is very short distance as well as for every class student has to stand in long line of iris scanner for marking presence [5].

III. TECHNOLOGICAL STANDARD

NFC stands for Near Field Communication. NFC is a short- range high frequency wireless communication technology that enables the exchange of data between devices. NFC is an upgrade of the existing proximity card standard (RFID) that combines the interface of a smart card and a reader into a single device. The main characteristic of NFC is that it is a wireless communication interface with a working distance limited to about 10 cm. [6]. The NFC card used in our system is the NDEF NXP NTAG213 chip which has an Operating frequency of 13.56 MHz and data transfer of 106 kbit/s. They are used in contact less payment systems and allow mobile payment replacing or supplementing systems such as credit cards and electronic ticket smart cards.

IV. PROPOSED METHODOLOGY

This project aims to develop an NFC Based Attendance System that will mark and monitor student's attendance without direct involvement of the faculty which would and uploads student attendance data directly to the Database. The steps in the proposed system are as follows (Refer Fig.1): -

- The professors register themselves by entering their details i.e. Name, ID number, department, semester and password into the mobile application which will be installed in their NFC enabled phones.
- To begin attendance process the professor will login into the application using their ID and

password. After login the teacher will input the subject and time and starts the NFC reader page.

- The professor then passes the mobile phone to every student to mark their attendance.
- Students have to tap their NFC card on the lecture's smart phone. On tapping the NFC card to the mobile phone the application reads the number and sends this number to the database.
- In the database the student's attendance is recorded for that respective lecture.
- At the end of the lecture the teacher will stop the NFC reader and close the application.
- A website is also created in the proposed system where teachers login using their ID number and password to add, update and delete attendance of a student.
- Students can also create and account to view their attendance for a particular lecture. This way the students and professors can keep a track of the student's attendance.

Thus in the proposed system we will try to minimize most of the flaws of the existing system. It will provide real time tracking of attendance of the students and save time wasted on taking attendance manually.
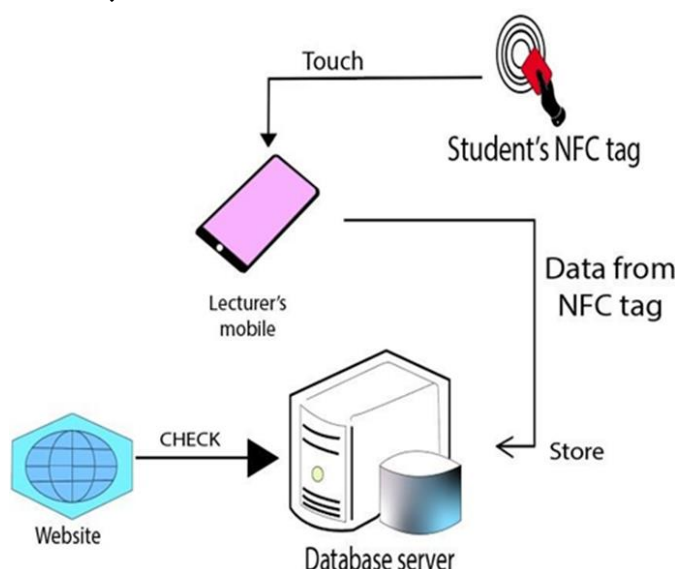
Fig. 1 System Design

V. ADVANTAGES

- There is no paper trail as the entire system is automated.
- Attendance will be directly stored in the database.
- All records are stored digitally.
- There are no manual functions in the proposed system so margin of error is low.
- Being low cost efficient its usage is more convenient.
- Efficiency in tracking and monitoring student's attendance will be excellent and accurate.
- Hardware required is minimal.
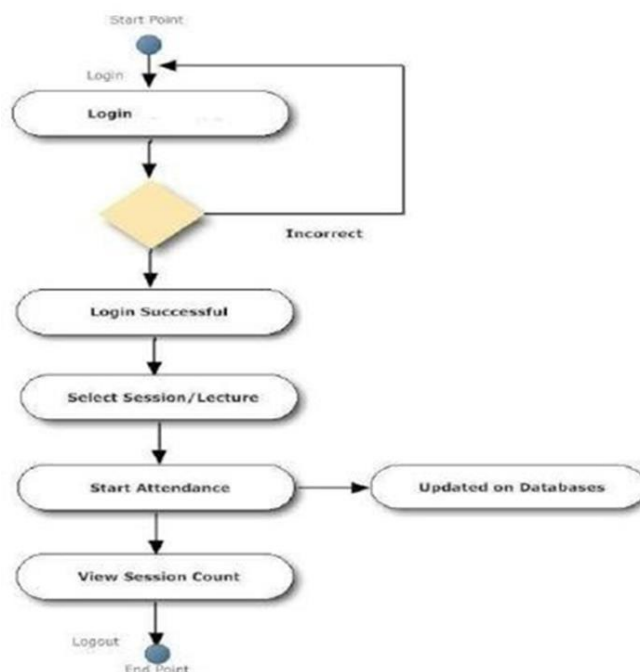
Fig.2 Faculty Activity Diagram

VI. IMPLEMENTATION

This project focuses on two aspects. The first aspect developed in this project is the Mobile application and the second aspect is the website.

A. Mobile Application

The mobile application is specifically designed and developed for the faculty members. The mobile application is developed using flutter in android studio.
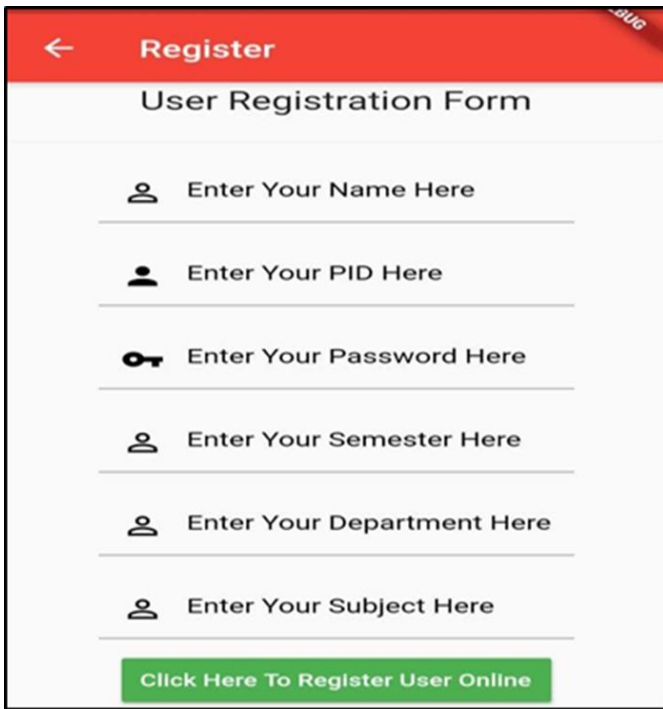
Fig. 3 Faculty Registration Page.

The faculty members have to first register themselves by clicking the registration page and filling in their respective details.
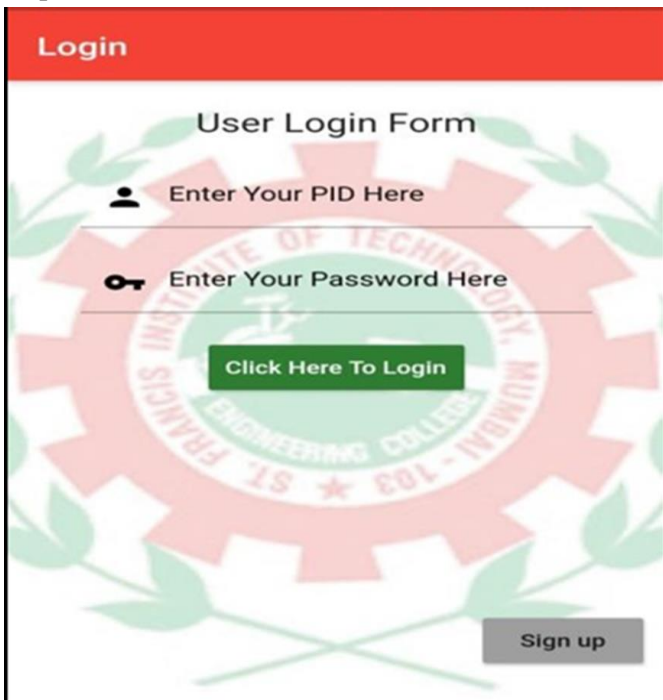
Fig.4 Faculty Login Page

After completing their registration the faculty can then login using their personal ID (PID) and their password.
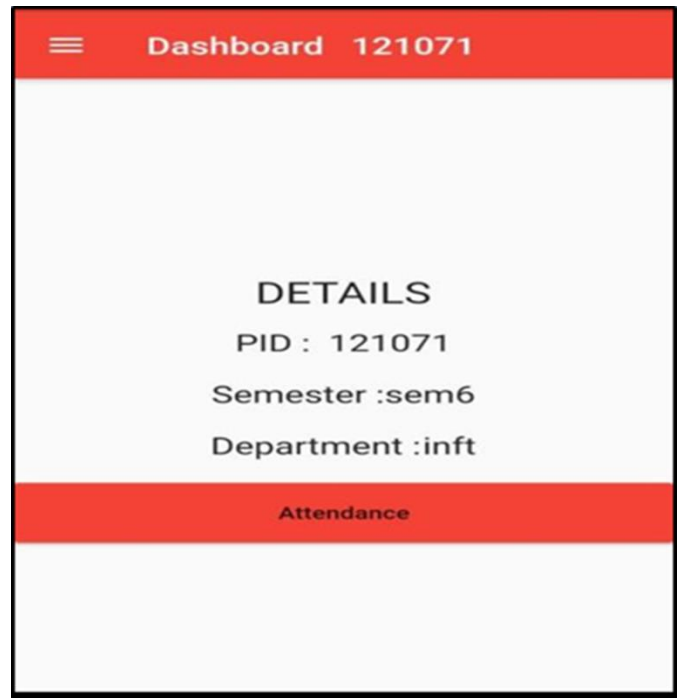
Fig.5 Dashboard

On successful login the faculty will be redirected to the dashboard (refer Fig.5) where the faculty page will be displayed. There is also an attendance button which when clicked will redirect you to the Drop down menu (refer Fig.5).
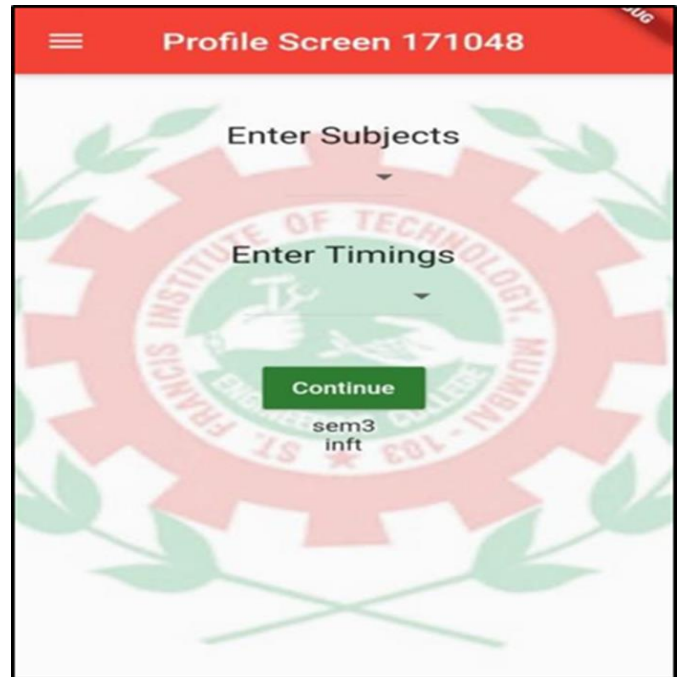
Fig.6 Drop Down Menu

The faculty then has to choose the subject and the timing of the lecture which he/she wants to take attendance for. The semester and department which

the user had filled in the registration page (refer Fig.3) is displayed below the continue button (refer Fig.6).
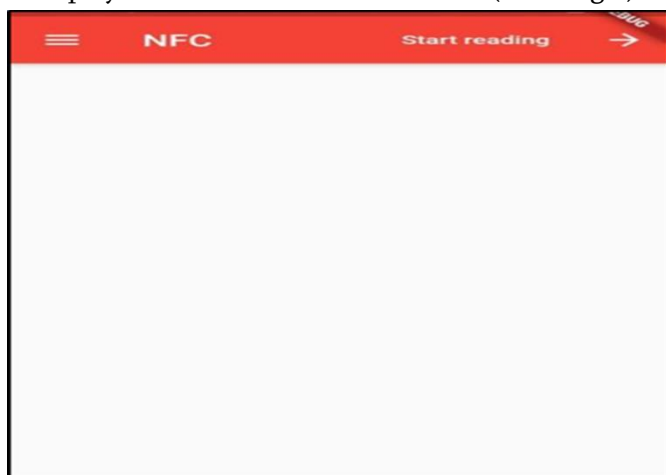
Fig.7 NFC Reader Page

The user is then directed to the NFC reader page(refer Fig.7)

On pressing the "start reading" button the NFC reader is enabled and the faculty can then pass around the phone for the students to tap their NFC ID cards onto the mobile phone.

Fig.8 Students Attendance gets Recorded

On tapping the NFC card onto the phone the attendance gets marked and their details i.e. their name and personal ID gets displayed.
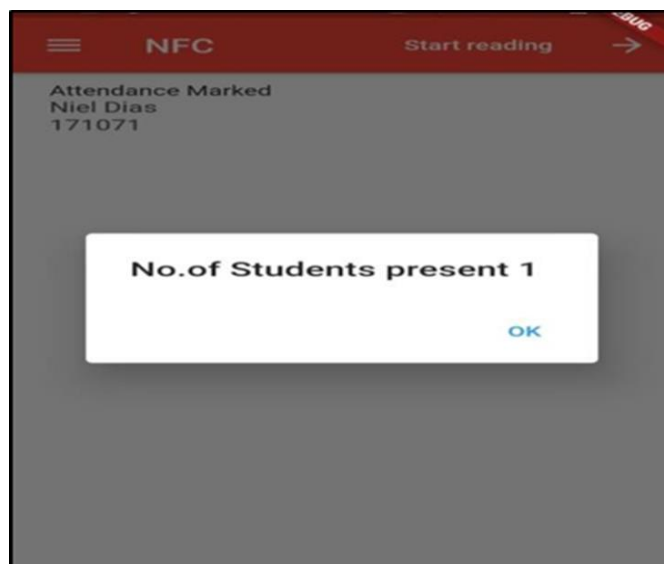
Fig.9 Number of Students Present

Once every students attendance gets marked the faculty can then click the "stop reading" button (refer Fig.8). A pop-up is displayed showing the number of students present for that lecture (refer Fig.9). This data is saved in the database and can be viewed via the website.

B. Website

The website is developed for both the students and the faculty. The faculty can view the attendance for a particular student or for all the students who were present for a particular lecture any students attendance; the students can only view their attendance.
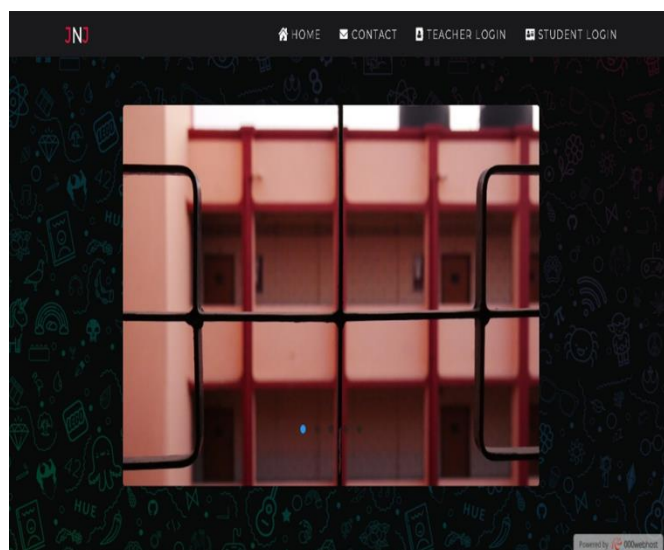
Fig.10. Home page

The home page is common for both the faculty as well as the students. The faculty and students can both login using the Login button in the home page (refer Fig. 10).
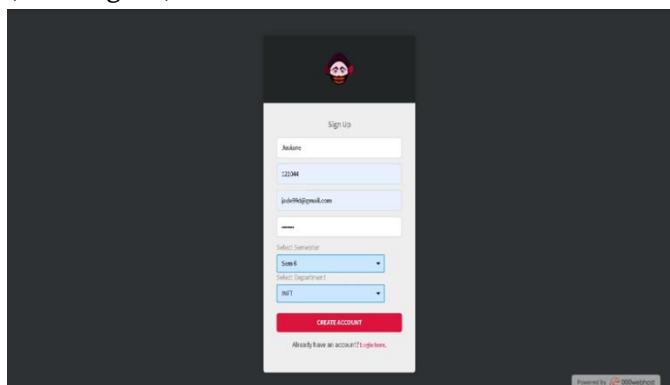
Fig.11 Faculty Registration

The faculty registration can be done either on the app (refer Fig.3) or on the website (refer Fig.11).
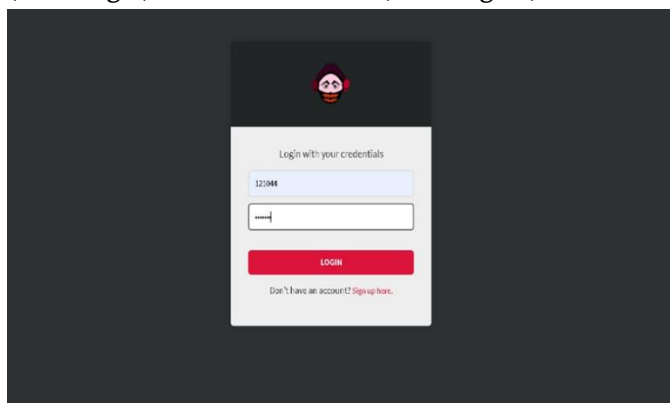
Fig.12 Login

The login module is same for faculty and students (refer Fig.12).
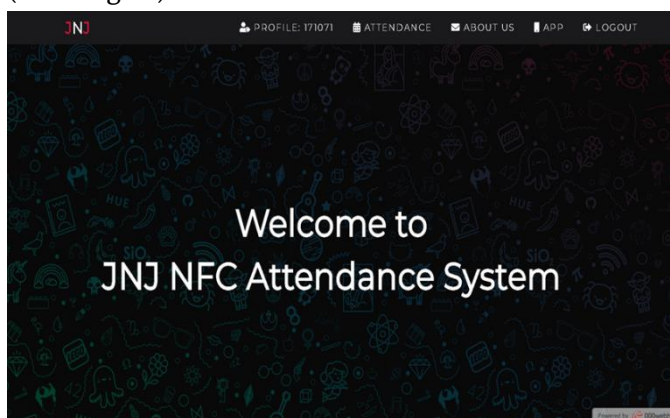
Fig.13 Dashboard Page

On successful login the user will be redirected to dashboard page and a session is created with their Personal ID (refer Fig.13).
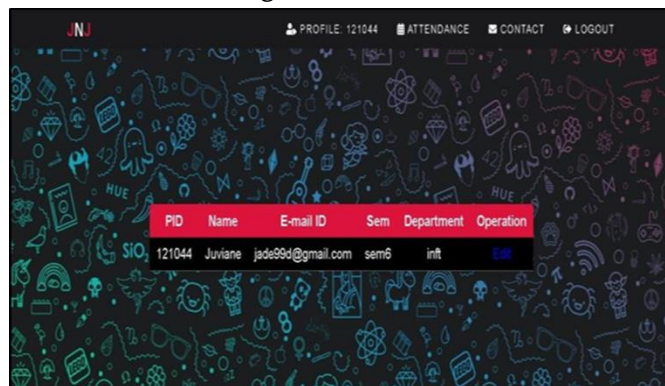
Fig.14 Edit Faculty Details

The faculty can also make changes to their details such as the semester and department by clicking on the edit operation (refer Fig.14).
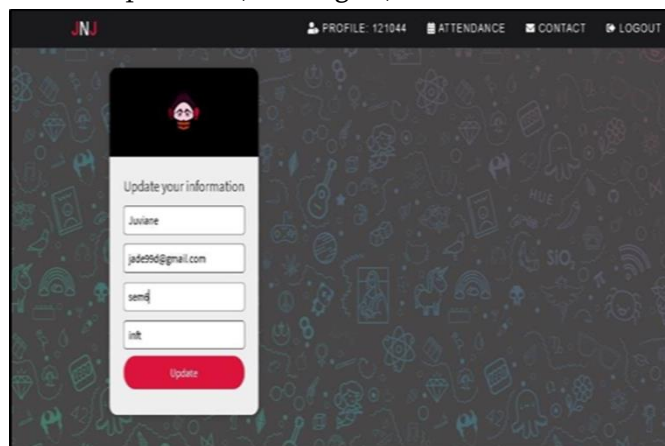
Fig.15 Update Details

The faculty will be redirected to this page after clicking the edit button (refer Fig.15) and can make the required changes.
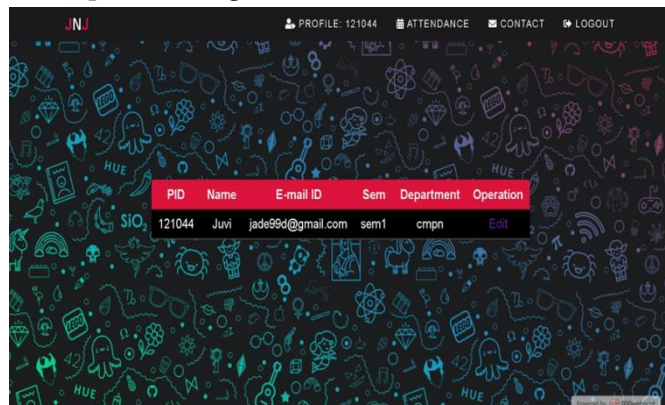
Fig. 16 Updated Records

Once the user clicks on the update button (refer Fig. 15) the changes will be displayed (refer Fig.16).
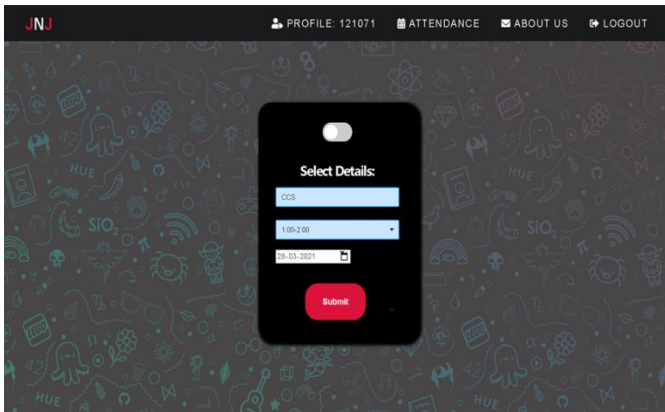
Fig. 17 Checking Attendance of all the students for a lecture

To check the attendance for all the students, the faculty has to enter the lecture,timing and the date for that particular lecture (refer Fig. 17).
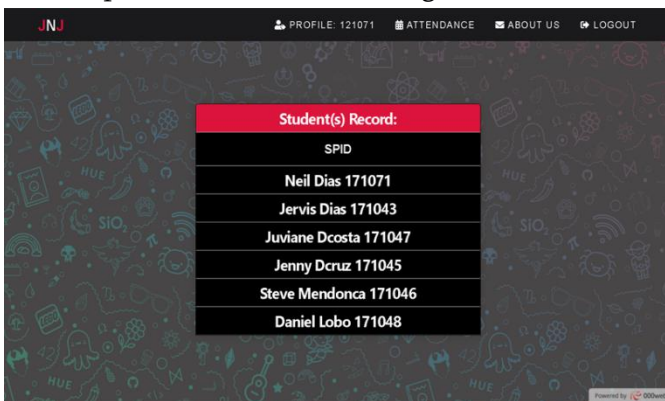
Fig. 18 Students Present

This will show the students present along with their Personal ID (PID).
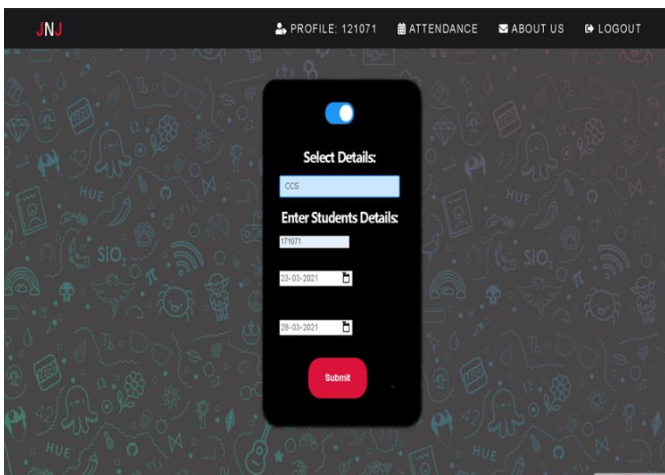
Fig. 19 Checking a Particular Student's Attendance

If the faculty member wants to check the attendance of a particular student, then they must enter the students Personal ID (PID) the subject for which they want to check the attendance, time of that lecture and also the range of the dates from when the faculty wants to check the attendance for.
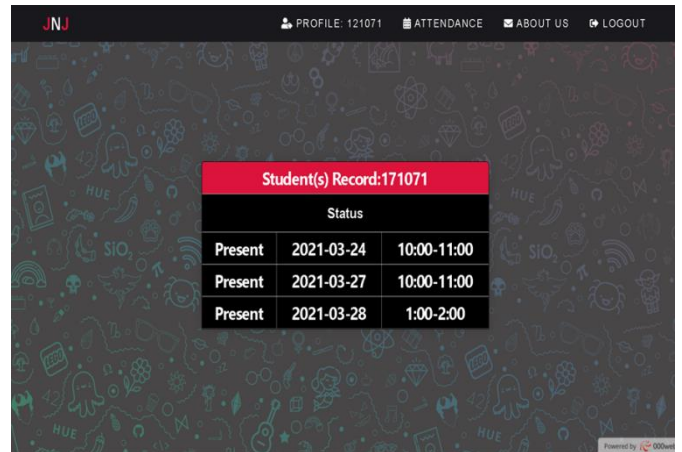
Fig. 20 Student Present

If the student was present then the status will show that the student was present (refer Fig. 20).
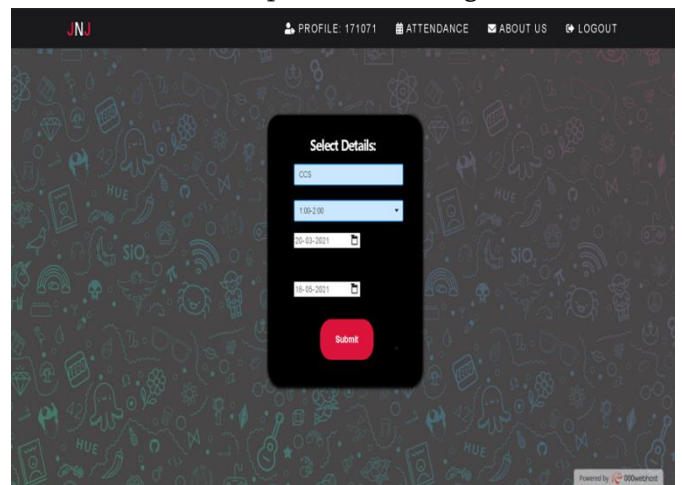
Fig. 21 Students Attendance

If a student wants to check their attendance for a lecture, then he/she needs to choose the subject for which they want to see their attendance for from the drop down menu, the time of the lecture and the dates from which they want to check their attendance from (refer Fig.21).
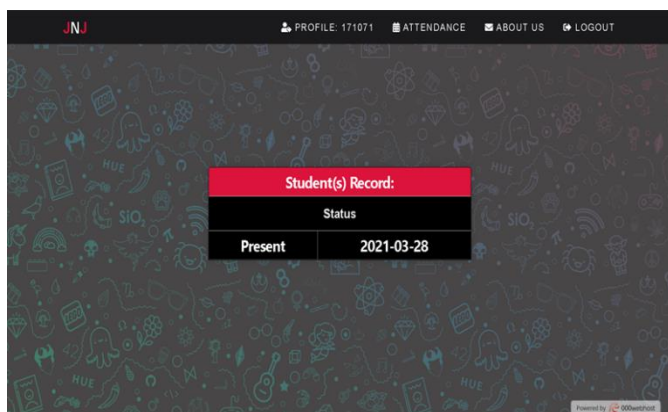
Fig. 22 Student Present for Lecture

If the student was present for the lecture then the student record will display present (refer Fig. 22).

VII.   FUTURE SCOPE

- The locking system will be replaced by personalized NFC enabled devices and NFC tags as door locks. Thus NFC will act as a farther step towards the world of automatic devices.
- It can also be used in shopping mall to scan a product and add it into a cart.
- It can be used to track objects using NFC tags.
- It may be used as attendance system for any educational institution.

VIII.   CONCLUSION

The NFC system is flexible and very easy to use.Additional security can be added into the cards so as the students cannot write on the NFC card. The NFC cards stored the data such as the PID of the students.This system can be put to use at any educational level and it can replace student Identity cards. As demonstrated, students, can use these cards for ease of attendance. NFC technology is ever growing, and the time has come for us to harness for ourselves, its potential and abilities. The main aim of this project was to demonstrate usage of NFC technology and build a simple attendance system based on it.

IX.   REFERENCES

[1]. Aparna Behara and M.V. Raghunadh, "Real Time Face Recognition System for Time and Attendance Applications", International Journal of Electrical, Electronics and Data Communication, Volume- 1, Issue- 4, 2013.

[2]. Research Note, Automating Time and Attendance: Low Hanging ROI, Proceeding in Nucleus Research, January 2008.

[3]. BISAM- BIS Attendance System by BIS Software ices PVT Limited. Available:http://www.softwarehouse.co/school-attendance-brochure.pdf

[4]. Vishal Bhalla, Tapodhan Singla, Ankit Gahlot and Vijay Gupta, "Bluetooth Based Attendance Management System", International Journal of Innovations in Engineering and Technology (IJIET) Vol. 3 Issue 1 October 2013.

[5]. Seifedine Kadry and Mohamad Smaili, "Wireless attendance management system based on iris Recognition", Scientific Research and Essays Vol. 5(12), pp. 1428-1435, ISSN 1992-2248, 18 June 2010.

[6]. Haselsteiner, Ernst, and Klemens Breitfuß. "Security in near field communication (NFC)." Workshop on RFID security. Vol. 517. No. 517. sn, 2006.

[7]. Prof. Krishna Tripathi, Bijoy Kunnappillil, Rohit Meshram, Vicky Bhoir," NFC- Based Attendance System" International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016.

[8]. Srivastava, Prerna, and Siddhartha Tiwari. "Review Paper of Design and Implementation of an Intelligent Biometric Attendance System using IoT.", International Research Journal of Engineering and Technology (IRJET), Vol. 7, Issue 01,ISSN 2395-0056, January 2020.

Real time moving Object Detection and Tracking using Machine Learning and Computer Vision

Pakruddin B[1], Farhanullah Shariff I[2], K Mohamed Tajammul[3], Suhail Khan[4], Syed Gibran Ahmed[5]

[1]Assistant Professor, Acharya Institute of Technology, Bangalore, Karnataka, India

[2]Full Stack Web Developer, Cognizant, Bangalore, Karnataka, India

[3]Full Stack Web Developer, Lunches, Bangalore, Karnataka, India

[4]Information Security Engineer, DXC Technologies, Bangalore, Karnataka, India

[5]QA Automation Engineer, Securiti.ai, Bangalore, Karnataka, India

ABSTRACT

The drone technology has been rising rapidly in the past few years, hence giving rise to better implementation of surveillance, monitoring, entertainment and research projects. In this context, our project presents an enhanced technique for detection of life (humans and animals) during natural calamities (like floods or forest fire) or to detect and rescue people when they go off-track in forests or deserts. It can likewise be utilized in observation purposes as unmanned flying vehicles (UAV) to catch still pictures and video to assemble data about explicit targets, which may be people, gatherings or conditions even in terrains which have been cannot be accessed by man himself.

Keywords — Object detection, object tracking, computer vision, machine learning, drone, UAV, YOLO, IoT, CNN.

I. INTRODUCTION

OBJECTIVES OF THE PAPER ARE AS FOLLOWS

1. Finding a way in which object detection can be done with minimal computational power on a raspberry pi which is connected to a battery and placed over a drone along with a pi camera.
2. Finding a solution to deal with loss of data over live streaming as internet connection maybe lost in deserted areas.

The popularity of UAVs (unmanned aerial vehicles) is increasing with the latest technologies with a large scope of applications. UAVs are used for surveillance, shipping, aerial photography, precision agriculture, object detection and much more. For example, reaching out hazardous environments or abandon areas. The UAVs give a clear scenario of the location where it is flying through camera mounted on it. The UAVs are commonly known as drone which are capable to fly without any human pilot on it, by using embedded systems it can fly autonomously. Drones these days are more compelling as to those of previous ones.

Drone with mini computers mounted on it such as raspberry pi, snapdragon processor series and various comparable boards increments the capacity of the drone. Motion detection and object detection algorithms provides better and faster control on the drone for autonomous flight controls. Mostly

commercially available drones do not have these many features in them.

Object detection is a computer vision method for finding occasions of articles in pictures or recordings. There are many techniques that can be used for object detection such as machine learning, deep learning, image segmentation, blob analysis, feature-based object detection and many more which are discussed in Section III. The most commonly used techniques for object detection are machine learning and deep learning and the others include image segmentation, blob analysis, feature-based object detection. But for Object Detection we will be using Haar feature-based cascade classifiers as it's lightweight and runs faster on the raspberry pi when compared to others.

The cameras used can be both color camera (ex: 3DR drone camera [9]) or an infrared camera [7], also called as the thermographic camera (forms a heat zone image using infrared radiation) and then the different objects are classified based on different temperature.

All the hardware and software under flight controllers such as pixhawk or ArduPilot project is open-source. As per his needs and special requirements, a user can modify the auto-pilot. Many flight controllers can be used in many special purposes and applications like in Obstacle Detection and Avoidance of obstacles. More about this is discussed in section IV.

Telemetry is generally defined as the automated measurement and wireless transmission of data from remote sources. A lot of drones usually have a telemetry attached to it. Using opensource based ground station software like MAVlink, you can even fly your aircraft with a joystick via your computer and update settings live. This is how the system proposed in this paper will be getting the live video output.

A GPS module is fit for getting data from GPS satellites and afterward to ascertain the gadget's land position utilizing appropriate programming and software, the gadget may show the position on a map, and it might offer directions.

GPS stands for Global Positioning System. Using this module, the proposed system will get the position of the human that has been detected by the camera that has been incorporated to the quad-copter.

The rest of the sections of the presented paper are separated as follows. Section II discusses the related work. Section III presents the algorithms surveyed which is followed by the flight controllers surveyed. Section V discusses the problem statement and Section 6 discusses its objectives. Tools and techniques and implementation are discussed in following sections i.e. section VII and section VIII. At the end we have experimental results and conclusion as section IX and section X respectively. References used are given at the end of the paper.

II. LITERATURE SURVEY

S. Yong and Y. Yeong [1] have proposed a method for forest surveillance which uses a large camera range covered which makes drone the perfectly required tool which can be used for much more advanced tasks. Much more advanced tasks require advanced algorithms, hence they have used a deep neural network architecture MobileNet and Single Shot Director (SSD) as the object recognition model.

A. Sarkale, K. Shah, A. Chaudhary and T. Nagarhalli, have [8] proposed an idea with the deep learning algorithm Faster RCNN as it's more efficient, accurate and the system can be implemented with speed and also because Faster RCNN can be used both for feature extraction and classification. They have four steps in the process, i.e. Gray Scale conversion, characteristic extraction, training and then classification.

S. Yong, A. L. W. Chung, W. K. Yeap and P. Sallis [9] have proposed an idea that incorporates motion detection algorithms along with commercial drones that can be helpful to tackle many problems like environmental changes, objects in motion and many

more similar applications. To identify the objects in motion, frame differentiating is added along with the algorithm which makes it much more for motion detection.

[12] X. Zhao, F. Pu, Z. Wang, H. Chen and Z. Xu have proposed a system that uses a UAV which has been incorporated with a monocular camera, inertial measurement units (IMUs) sensors and a GPS receiver. YOLOv3 is the algorithm that has been proposed to be used along with the system. The drone is a high end drone built on the platform. DJI M100 on which a monocular camera and a microcomputer Jetson TX1 are attached.

Getting a precise and consistence detection of a moving object is challenging, the complexity in detecting an object increases due the dynamic behavior of the background and the object compared to a static background. [14] To overcome such problem a combination of optical flow and gaussian mixed is implemented to provide reliable tracking. Having a reliable tracking system enables better detecting the object of interest, depicting the scenario where the drone could be used in forest, urban and deserted environments which include dynamic elements out of which the moving objects have to be detected using the method presented by the author [14]

called "pixel motion process", where successive pixels from each frame are analysed and mapped between every two frames based on this computation motion of an object is predicated for tracking this helps in generating real time computation data for reliable tracking. Unlike the aerial surveillance and CCTV cameras where the foreground is in motion compared to their background the subject can be easily traced therefore the key problem in detecting a subject in motion in such live feed is object detection with fast and reliable result with minimal delay since the operation is carried out real time.

Before getting into object detection and tracking we must understand that tracking is a method involves the rough calculation of path to identify and judge

changes that an object imposes in each successive frame from the live video feed. Kalman filter along with Camshift method is used to get precise tracking and detection even for a video with low video quality with noise and distortion [15][16]. But these methods were designed for a single object detection, since the aerial video may include more than one person a combination of hybrid method along with optical flow can be used to get desired result. Camshift is an upgraded version of Mean-Shift method [15] which can detect rigid and non-rigid objects which could be used to distinguish subjects for object detection. The common methods include.

1. Prediction 2. Correction 3.Association

In case of optical flow pixels are analysed form a sequence of frames building a motion sequence which is used to compare the background motion combined with Kalman's filter from both methods presented by authors [15][16] the object path can be tracked effectively.

III. PROBLEM STATEMENT

From the above literature survey that we have performed, the following problem statement can be stated.

Let's consider a scenario where a person is lost in a forest or a deserted place. Now how are drones useful in such scenarios. Drones can be used with attachments such as cameras and GPS.

Live streaming from drone cameras are common and its position can also be tracked quite easily. But live drone cameras send their live feed using internet. There's hardly a signal in remote areas such as forests and deserts. Live streaming is interrupted. The chances of finding the person in such cases is only as good as the internet connection. Yes, there are systems that record the footage and then the footage can be reviewed after the drone has arrived to the ground station (It will only come back if it at least has a stable GPS connection).

Now, most of the drones available in the commercial space are not autonomous. As the autonomous nature and the functions of the drones increases, the cost of the drones also increases and they can go up to Lakhs. There are still various voids that need to be filled in such departments. Example, as mentioned above, if a person is lost in a desert or a forest, and if we are using drones to detect where the person is, then one drone may not be sufficient to track down the person. Depending on the search area, that is usually large in cases of deserts and forests, many drones may have to be deployed in order for the search to be completed. There are a lot more constraints such as the battery of the single drone which is being used may not live up for a long time.

As mentioned above, if there's a loss of connection, then some of the footage can be lost, hence reducing chances of the person being found. What happens if the footage is recorded and then reviewed after the arrival of the drone at the ground station? It may take a long time to review and check every frame and detail. There could also be a chance that the person has moved away from that particular place where he was seen in the footage.

IV. ALGORITHMS SURVEYED

YOLO (you only look once): YOLO is a very quick continuous multi object recognition algorithm. The calculation applies a neural system to a whole picture. The system partitions the picture into an S x S matrix and comes up with bouncing boxes, which are boxes drawn around pictures and anticipated probabilities for every one of these regions. But to run this, a good computing power is required. The algorithm can run on the raspberry pi 3 but there will be an everlasting lag. Our aim is to run an algorithm

SVM: Support vector machines. A quick and reliable classification algorithm that performs very well with a little amount of data. It is a supervised machine learning model. It uses classification algorithms for

two-group classification problems. If labelled training data for each category is given to an SVM, it will be capable to categorize the new data.

RCNN: The RCNN calculation proposes a lot of boxes in the image and checks if any of these boxes contain any article. RCNN utilizes selective search to extract these boxes from an image (these containers are called districts). But, it is slow. It is quite hard to train as we have to train CNN, SVM and bounding box regressor separately. You have to save every feature map of each region proposal and that needs a lot of memory.

Fast RCNN and Faster RCNN: Keeping in mind that we have to run the algorithm using the minimum of RAM, we are using a raspberry pi 3 which has a RAM of 1GB. We cannot run Fast RCNN and Faster RCNN on such low RAM. So, don't consider these algorithms. Haar cascade classifier takes the positive image (image in which the object is present) and negative image (image in which object is not there) of the object and trains the model for detecting the object. It extracts the features of the object from its positive and negative images and classifies the features into different phases. While detecting the object Haar classifier compares the frame features with the stored feature phases of the trained model. It does it in steps is the first phase of classified feature fail to match it discards the detection model else it continues comparing with different phases of features.

So in the proposed project, we have decided to use the Haar cascade classifier as it is lightweight and runs freely occupying very little RAM on the Raspberry pi 3.

TABLE 1: Different Methods used

Method/CNN Model	mAP(%)	Speed (FPS)
DPM	33.5	0.07
RCNN	65.7	0.05
Fast—RCNN	69.8	0.50
Faster-RCNN	73.1	5.00
YOLO	63.2	45.00
SSD 300x300	73.9	46.00
SSD 500x500	76.8	19.00
YOLOv2 416x416	76.8	67.00
YOLOv2 544x544	78.2	40.00

V. FLIGHT CONTROLLERS SURVEYED

1) KK 2.1

A KK 2.1 is one of the basic flight controllers. It is very easy to use because of it's simple architecture. Although it does not support automation flight, it is very commonly used for manual drone flights. Also, it doesn't have any inbuilt sensors, which is one of the disadvantage of the flight controller.

2) ARDUPILOT

The ArduPilot Project gives a propelled, full-included and dependable open source autopilot software system. The first ArduPilot open code archive was made in 2009 - from that point forward it has been created by a group of differing proficient designers, scholastics, PC researchers, and different individuals from the worldwide network. It is equipped for controlling practically any vehicle framework comprehensible: conventional and VTOL, helicopters, airplanes, sailboats, gliders, multi-rotors, ground vehicles, powered boats, submarines and even Balance-Bots. The bolstered vehicle types every now and again grow as use cases develop for new and novel stages.

3) PIXHAWK

The Pixhawk flight controller is an open source hardware and is very powerful when compared to the ArduPilot flight controller which is why it became much more popular for DIY drone builders. The ArduPilot(APM) uses a 8-bit processor whereas a pixhawk uses a 32-bit processor. There are many ports in an pixhawk which makes it easy and efficient to communicate with the computer systems and other external devices. The pixhawk also comes with inbuilt sensors like GPS, magnetometer, LiPo power module etc.

4) PIXHAWK CUBE

The pixhawk cube is an advanced version of the original pixhawk flight controller. It is sometimes referred to as pixhawk 2. One of the main difference between the original pixhawk and the pixhawk cube is the functionality and affordability factors, the original pixhawk comes is much affordable when compared to the pixhawk cube. If the project demands for more online support then pixhawk cube is probably the right choice.

5) NAVIO2

Navio2 is one of the most reliable flight controllers in the current time. With other flight controllers we need to append raspberry pi externally for additional computation, but Navio2 flight controller acts as a shield which can just be mounted on top of the raspberry pi. Navio2 comes with a free Debain OS. Also Navio2 is one of the first flight controller to support full Linux distribution OS.

In this proposed system we are using APM 2.8(ArduPilot) flight controller as it is reliable, costs lesser when compared to other flight controllers and also provides autonomous control over the UAVs.

VI. TOOLS AND TECHNIQUES

RASPBERRY PI:

The Raspberry Pi is a progression of small single-board PCs created in the United Kingdom by the Raspberry Pi Foundation. The Raspberry Pi is a credit-card sized computer that can be connected to a monitor or a TV screen, and uses a standard console and mouse. It can do everything that you'd anticipate that a personal computer should do and has the ability to interact with the outside world.

Operating System – Raspberry Pi has the capability to run Operating Systems like Linux, Windows, Android etc.

Storage – The storage uses microSDHC slot.

Memory – 1GB, 2GB and 4GB RAM.

CPU – 1.5GHz with 32/64-bit quad-core processor.

PI CAM:

The Raspberry Pi camera is a high quality image sensor module designed for connecting onboard for Raspberry Pi. The camera module can be used to capture still images and HD videos. It can capture still images up to 3280 pixels and supports 640p, 720p and 1080p videos. In our project the PI cam takes in the live feed and sends it to the Raspberry Pi.

DRONE HARDWARE PARTS:

1) ESC: ESC stands for Electronic Speed Controller. It is used to convert the PWM (Pulse Width Modulation) signal from the flight controller to run the brushless motor by supplying adequate amount of electrical power.

2) MOTORS: A motor is a bunch of coils (electromagnets) which are connected together in pairs. The motors are associated with the propellers which makes the propellers to rotate and produce thrust to lift the drone against the gravity. The motors can rotate in both clockwise and anti-clockwise direction.

3) PROPELLERS: Propellers are fixed on top of the motors. The propellers sizes vary from 3inch to 23inch. The lower propeller size requires higher motor speed and vice versa. In our project the size of the propeller used is 6inch and the speed of the motor is 1000kv.

4) FRAME: A frame is the backbone of a drone which holds all the components of the drone in an efficient manner so that the weight of the components is equally distributed in the drone. The frame size varies from 120mm to 500mm. The frame size used for our project is 450mm.

5) TRANSMITTER: A transmitter or a remote is a drone controlling device which transmits the

commands to the receiver with the help of radio wave signals. The project uses Flysky fs-i6 model whose frequency ranges from 2.4055 to 2.475 GHz.

6) RECEIVER: The Receiver is mounted on top of the drone and is used to receive commands from the transmitter. The receiver further forwards the commands to the flight controller. The Receiver has a pair of antennas which is used to receive the signals from the transmitter.

FLIGHT CONTROLLER:

A flight controller is circuit board responsible to direct the RPM (Revolutions per Minute) of each motor. The command received from the receiver is fed into the flight controller which dictates how to manipulate the motors accordingly. The flight controllers also consists of sensors like Gyroscope, Barometer and GPS. A Gyroscope is used for orientation, Barometer is used for holding the drone at some altitudes and GPS is used to trace the location of the drone. There are different types of flight controllers namely Pixhawk, ArduPilot, KK 2.1, etc. used for various purposes accordingly. In our project we used ArduPilot flight controller which is best suited for multi-rotors.
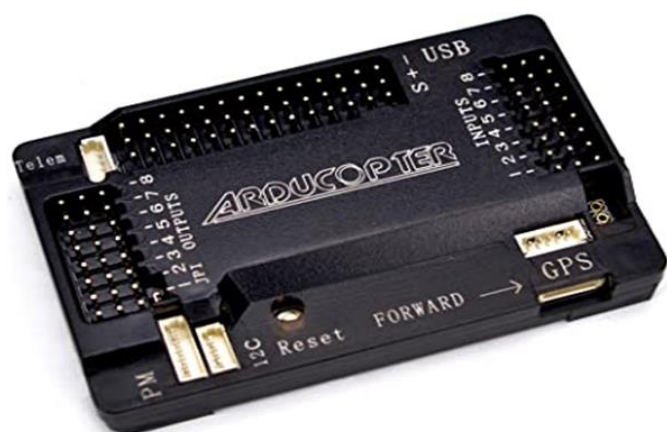
ALGORITHM USED:

We find a large number of algorithms available for the purpose of object detection. In the available list of algorithms for object detection we see algorithms which are very accurate require high performance computation machine for training the modules for object detection, also we see few algorithms which do not require much computation power but are not much accurate for detecting an object. In our project we are using Raspberry pi 3 for the purpose of object detection which is a good compact computer on single board (but not the best). We chose raspberry pi as it is light weight requires less battery power to run and can be fitted on a drone easily. As a single board computer, raspberry pi gives a decent performance but it can't be used for algorithms such as RCNN, faster RCNN or YOLO as they use tensorflow APIs which are heavy for our raspberry pi. After research on the algorithms available, considering the computational power of raspberry pi we came in to conclusion that Haar cascade object detection works fine for the raspberry pi and gives a decent accurate result. Haar cascade object detection algorithm requires many positive images (images with the object) and negative images (images without the object) to train the classifier module. To train the classifier module the features are extracted from the images by subtracting sum of pixels under the white rectangle from sum of pixels under the black rectangle (as shown in the figure). The features are assembled into various phases of classifiers and applied individually (Usually the initial stages will contain less features). While detecting an object, if the frame does not succeed the first stage, then it is discarded. The remaining features are not considered. The process is continued if it passes the next stages. If the frame passes all the stages, it detects the object for which the classifier model was trained.

VII.   IMPLEMENTATION

1) PROPOSED SOLUTION

The idea of object detection is to identify and target the desired object in the frame , in a typical setup the target is identified after the footage us taken this causes delay in identifying the object , by implementing objection detection real time various

---

application were possible , which limited due to constrains. During a calamity or a disaster it is a tough and tedious task to identifying people that are trapped due disaster, by using a drone we can get an aerial view of the desired area without the risk of out, but identifying the desired target it not easy with naked eye, the task can be made easier and accurate with the help of machine learning and computer vision. The first step in achieving this by training and developing a model to detect the target by proving it with lots of trading data for example if the target to be identified is a person then the training data consists of images consisting of a person by doing this step with the help of machine learning. A model is developed which can be used by the drone camera to identify the desired target. The whole setup consists of a drone attached with a camera along with radio control and raspberry pi, the goal here is to provide real time object detection by taking advantage of the setup, a web client and a server is created to stream the video and perform object detection respectively. By utilising the processing on the raspberry pi the objects are from the camera output are detection in real time and sent to the end user, therefore providing real time detection of objects, this is an important factor for many application, once the objects are identified an email is send to the user highlighting the target in the give area that is detected by the drone. The drone can be operated with radio control with long range and the video feed is transmitted via internet which provides flexibility and controls in remote area
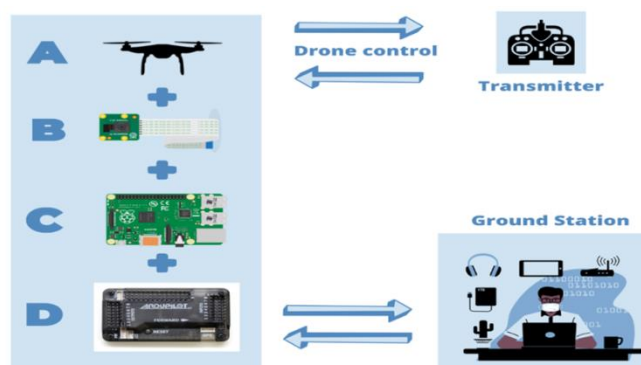
2) ARCHITECTURE

The architecture consists of a drone carrying essential hardware and a ground station , the drone consists of [A]frame mounted with [B] 5 MP pi camera and [C]raspberry pi which is controlled by [D]APM flight controller , the ground station consists of radio controller along with a device such as mobile or laptop where the video and flight data can be streamed, the idea of the this setup is to allow real time object detection in a drone, flight control and object detection is handled separately by specific hardware ,the dynamic controls and stabilization of drone is carried out by APM flight controller giving flexibility for processing objection detection. Raspberry pi along with pi camera is used to provide computer vision and objection detection for the drone, the video feed from the camera is processed by raspberry pi with the help of machine learning and image processing, since the processing is taking place on board as the data is produced this gives rise to real time object detection.

Raspberry pi also handles the task of running a server which is used to transmit the video feed to the user. Multiple device can be used to see the live video feed from the drone along with objection detection. This is done by creating a web client and server interface so that the video stream can be sent via an internet connection. The ground station is used to monitor the flight data consisting of location, altitude, angle etc. A radio transmitter is used to control the drone each of this instruction is received by the flight controller through a receiver. It determines the amount of power to be power to be transmitted to each motor to keep the flight stable. The ground station can also be used to view the video stream. Therefore, this system allows the operation to be carried out with real time.

3) OBJECT DETECTION

Object detection is a computer vision method for finding occasions of articles in pictures or recordings. There can be different classes into which objects that are identified can be grouped into, such as animals, humans, plants, cars, buildings, etc. Domains of object

detection include face detection, anomaly detection, and pedestrian detection. For object detection in our project we have used a credit card sized computer raspberry pi 3. It is coded to identify humans, the language use is python, and it is much easy and reliable language used for ML and AI coding. The library of predefined functions made python an ideal language for application such as image detection. The libraries of OpenCV are included for processing the image. The OpenCV includes Haar classifier, it is the simplest and most effective method for object detection. In Haar classifier the module for object detection is trained by using the negative and positive images of the human body. Haar classifier then classifies the features of the human body from the images. The pi camera module attached to the raspberry pi 3 takes the live feed, after which the Haar classifier compares the frames from the live feed of the pi camera with the features of the human body extracted while training the module. The features are assembled into various phases of classifiers and applied individually. While detecting a human, if the frame does not succeed the first stage of matching the features, then it is discarded. The remaining features are not considered. The process is continued if it passes the next stages. If the frame passes all the stages, it detects the human body and sends a mail of the still frame of the footage with human detected to the ground station.

VIII. EXPERIMENTAL RESULTS

Fig. 1 Object detection with accuracy 70-80%

Fig. 2 Multiple Object detection

IX. CONCLUSION

In this paper we have proposed a light weight object detection and tracking method using Tensor flow lite model, running on a raspberry pi, attached to a drone. Real time performance in the drone can be achieved through our proposed solution. Different demands can be met as our design is easily extendable. In the proposed method, we examined how to track objects efficiently and the results have been analyzed through probabilistic approaches. Through the results, we prove that a simple algorithm in a raspberry pi attached to a drone can be of substantial and pronounced use.

X. REFERENCES

[1]. S. Yong and Y. Yeong, "Human Object Detection in Forest with Deep Learning based on Drone's Vision," 2018 4th International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2018,pp.1-5.doi:10.1109/ICCOINS.2018.8510564

[2]. Jianchao Zeng, A. Sayedelahl, M. F. Chouikha, E. T. Gilmore and P. D. Frazier, "Human detection in non-urban environment using infrared images," 2007 6th International Conference on Information, Communications & Signal Processing, Singapore, 2007, pp. 1-4. doi: 10.1109/ICICS.2007.4449646

[3]. A. Sarkale, K. Shah, A. Chaudhary and T. Nagarhalli, "An Innovative Machine Learning Approach for Object Detection and Recognition," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1008-1010. doi: 10.1109/ICICCT.2018.8473221

[4]. S. Yong, A. L. W. Chung, W. K. Yeap and P. Sallis, "Motion Detection Using Drone's Vision," 2017 Asia Modelling Symposium (AMS), Kota Kinabalu, 2017, pp. 108-112. doi:10.1109/AMS.2017.25

[5]. X. Zhao, F. Pu, Z. Wang, H. Chen and Z. Xu, "Detection, Tracking, and Geolocation of Moving Vehicle From UAV Using Monocular Camera," in IEEE Access, vol. 7, pp. 101160-101170, 2019. doi: 10.1109/ACCESS.2019.2929760

[6]. Y. Wang, Z. Zhang and Y. Wang, "Moving Object Detection in Aerial Video," 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, 2012, pp. 446-450. doi: 10.1109/ICMLA.2012.206

[7]. G. Swalaganata, Muniri and Y. Affriyenni, "Moving object tracking using hybrid method," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, 2018, pp. 607-611. doi: 10.1109/ICOIACT.2018.8350740

[8]. A. Kalyankar, S. Nema and U. Mahind, "Advance and Automatic Motion Detection, Prediction, Data Association with Object Tracking System," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1442-1444. doi: 10.1109/ICIRCA.2018.8596787

Anybody Can Dance Using Human Pose-Estimation and Transition

Preksha Shridhar[1], Surya G J[1], Chempavathy B[1]

[1]Department of Computer Science and engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

This the approach of using pose or stick figure as an intermediate for video-to-video translation using pose as an intermediate representation, we will be able to transfer the motion from the given original video to the target video. To transfer the motion, we need to extract the poses from the source video and use that to man and change the target video motion. We take two consecutive frames for temporarily rational video results and introduce a separate pipeline for more realistic face synthesis. It produces surprisingly compelling results. This motive to also provide a forensics tool for reliable synthetic content detection, which is able to distinguish video that is integrated by our system from real data. This approach is applied to a given source video of a professional dancer from which the motion is transferred to the target video of an amateur dancer to perform the standard moves.

Keywords — Open Pose, Pose Estimation, Normalisation, Key-Points, CNN.

I. INTRODUCTION

Consider the two videos, a source video of a person dancing ideally a professional for best result. We will be able to transfer that routine to another person called the target after collecting few minutes of training data from the target to perform standard moves. Implication of this technology are profound like autotune for dancing. No needed necessarily to dance well in order to create a video of ourself dancing.

To capture motion, mo-cap specialists place sensors all over an actor. These tracks and record their movement, allowing them to be mapped on computer screen in real time as a virtual skeleton. Animators then use these computer program to overlay information on top of the movements.

Our goal is to extract the position of each of the body parts of every single person appearing in the image with no more sensors than a digital camera using photos and videos captured.

We are going to design this model using the human estimation method. Open pose is the library that would help us get what we are looking for. This library consists of some neural network and some other functions. Neural network is converted to a format that TensorFlow understands. Later there is a need to make some post-Processing on this super tensor in order to get valuable information out it.

Open pose extracts feature from the image by using first few layers. Then the features are fed into parallel branches of convolutional network layers to extract the pose.
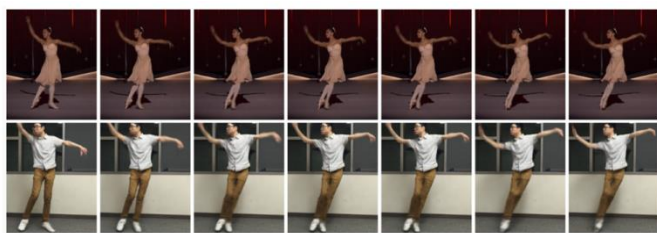
Figure 1. Motion Transfer from Source to Target.

II. PROBLEM DEFINITION

The top Figure 1 is the clip of the dancer i.e., Ballerina (source object) performance of a sequence of motions. The bottom row is the result of the algorithm that is implied as the output. It is matching to the frames of a different persons that is the target subject, seemingly performing the same motions. The twist here is that the target person never performed the same series of motions as the source, and indeed, knows nothing about ballet dance. He was actually filmed by performing a set of his own moves, without specific reference to the accurate actions of the source. And, also notice the above-mentioned figure, the source and the target are of different gender, have different physique and different clothes.

III. TECHNOLOGY

The technology we are going to use in this project is the one which can help us detect human posture and the motion capture technique and GAN.

Human pose estimation has become an important problem in the field of computer vision. Imagine being able to track person's every small movement and their details and do a bio-mechanical analysis in real time. The technology will have a huge implication. Pose estimation is basically predicting the body part or joint positions of a person from an image or video.

Motion capture is actually a descendance of one of the oldest animation techniques, which is known as rotoscoping. This is the process where the tracing over live action footage to make an animated film.

Motion capture is the result of the combination of rotoscoping with newer computer technology, where it allows people to use live footage as the basis for animation without having go through the capture.

A pose skeleton of human represents the orientation of a person in the graphical format. It is essentially a set of coordinates that is to be connected to describe the pose stick of the detected person. Each of these co-ordinates in the skeleton are known as a part or a joint or a key-point.

So, the goal of our machine learning model is to track these key-points in the video given as the input. With this method we are able to track an object or person in real-world space at an incredible granular level. This powerful capability opens up a wide range of possible application.

Some of the applications includes: Augmented reality, animation, gaming, robotics.

Pose estimation is one of the elegant applications of neural networks and is startingly accurate and sometimes, seems like something right out of science fiction.

Strong articulations, small and vaguely visible joint points, occlusions, clothing, background and lighting changes makes this approach a bit difficult problem. The classical pipeline has its limitations and pose estimation has been greatly reshaped by CNN.

Video is a series of set of images; hence it'll be broken down to number of frames. If the temporal features are not a part of pipeline, there's a possibility to apply a static pose estimation for each frame in a video.

GAN is the second technology that we will be using in this project. This consists of two networks that is generator and the discriminator. Generator synthesizes the given input to a fake image and discriminator discriminates between the fake and the real output. By this way both the network is trained to such a way that generator will produce a fine fake output and the discriminator will adapt to it.

IV. PROPOSED SYSTEM

This project is all about the human pose estimation and transition of the pose skeleton body. Our basic goal here is to extract the position of each of the body parts of every person appearing in the image or videos with no more sensors than a digital camera using photos and videos captured.

We are going to design this model using human estimation method. Open pose is the library that would help us get what we are looking for. This library consists of some neural network and some other functions. Neural network is converted to a format that TensorFlow understands. Later there is a need to make some post-processing on this super tensor in order to get valuable information out of it.

The Open pose consists of 3 phase CNN structure. So, with the help of pose2pose we create a pose2motion. This solves the quality and loss issues and simplifies the pose recognition. We were will also be able to clone the pose and copy it to a new dataset which would be an amateur dancer.

Pose2motion not only detects poses and copies but it'll copy even the motion of the object in real time.

We used pose estimation for pose2motion detection which will also help us resolve deep fake by detecting the synthetic content using reverse engineer method.

We will also be using GAN (generative adversarial network) which is a method to obtain image date from all the process.

Using all the above method we will be extracting the skeleton joint from each of the source and the target video. The skeleton joint from the source figure is then normalised to the joints of the target subject. It is the transformation of key points of source to target as it maybe varies in height and ankle, hence it is analysed and using a linear mapping it is mapped to the closest and farthest ankle position in both videos.

Then we will translate the skeleton figure by using generator as it synthesizes images in order to trick and deceive the discriminator. But eventually both will be trained in such a way to produce a more detailed image.

V. METHODOLOGY

The training pipeline is split into 3 steps: Pose Estimation, Pose Normalisation, Mapping normalised pose stick figure to the target subject.

Finding the video of the source, they needed to be encoded better way to use it's pretrained pose detector called open pose. We can accurately estimate all of the subject's multiple join co-ordinates. This is a convolutional neural network. A specific series of matrix operations that was optimised for pose Estimation by using common strategy of gradient descent. Take the co-ordinates and draw the characterization of the resulting Stick Pose figure by plotting the Key Points, extending the lines between connected joints.

The frames of images are obtained from the input source video.

First method is pre-processing, where Pre-processing is the method where there is converting of the image from [0,255] to [-1, 1].

The next operation is the usage of the neural network where it returns a tensor consisting of 57 matrices. This last operation is just a concatenation of two different tensors i.e. Heatmaps and PAFs part affinity fields.

Heatmap is a matrix that stores the confidence points, the network has that a certain pixel contains a certain part. There are 18(+1) heatmaps associated with each one of the parts and indexed as shown.

PAF are the matrices that gives information about the position and orientation of pairs. They come in couples for each part we have a PAF in 'x' direction and 'y' direction. 38PAF associated with each one of the pair and indexed.

The location of body is extracted from these 18 matrices.
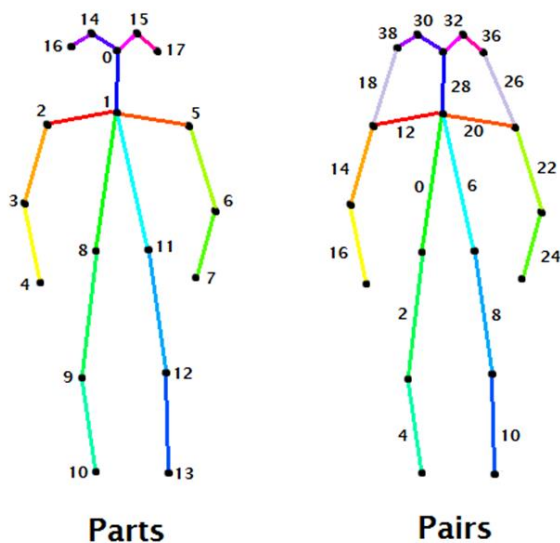
Figure 2. Parts and Pairs of Human Pose Estimation

The next step is detecting pairs of the image. We need to extract part locations out of a heatmap. NMS (non-maximum suppression algorithm) is applied to get those pixels.

From this algorithm the obtained non-zero pixels denote the location of the part candidates.

The next step is to form a bipartite graph. Each body parts should be connected to form pairs.

This graph is obtained by using the assignment algorithm which uses linear method to find the connection and to form a pair. These pairs will hence forth form a skeleton form of the human posture in the image.

Using PAF values, weaker links in the bipartite graphs are pruned. Through these steps the skeleton of human pose can be estimated and assigned to every human in the image.

Then the image translation is done to the trained model of the target subject. This is done by using generator method which helps in synthesizing the skeleton pose to the human body.

VI. EXPERIMENT

A. Data Set

Basically, we have two-part datasets: the first one being the single dancer video that is filmed by us and is used for training and then evaluating the model. The second one being the source video of a professional dancer which is taken from YouTube or any online medium.

The subjects from each dataset:

1) Test image which is the ground truth person images.
2) Corresponding stick pose figure of the images.
3) Corresponding face bound box co-ordinates.

The motion is transferred between the subjects of each datasets to provide the end result after training and transition.

B. Working Model

The given two datasets the subject is extracted from each of the datasets. These subjects of dataset is used as input to Generative adversarial network for training. The network is construction of two neural networks called the generator and the discriminator. The generator applies the series of transforms to the input image to produce the output image. The discriminators job is to then perform a binary classification trying to discern if output image is real or fake. That is if it's actually the target subject dancing or a fake version.

The structure of generator has encoder and decoder as showing in Figure 3. It takes the input image and to compress into a much smaller representation using series of encoders. These are operational blocks consisting of a convolution and an activation function. The idea is that by compressing it like this we can hopefully have a very prominent representation of data after the final encode layer.

The decode layer do exact opposite using operational blocks consisting of a deconvolution and an activation function and reverses the action of the encoder's layers. A performance improvement here is to

directly connect the encoder layers to decoder layers using skip connections. These skip connections give network the option of bypassing certain encoding and decoding parts. If it doesn't have a use for it. This specific type of encoder, decoder architecture is called a U-net. Meanwhile the discriminator has the job of using two images. One from target video and other from generated output and deciding if second image produced by generator or not. Convolution network structure looks similar to the encoder section of the generator but works a little differently. The output is an image where each pixel value represents how believable the corresponding section of the unknown image is.

There are two steps to train this network: Train the discriminator D and training the generator G. To train D, first G creates an output image. D looks at input target pair and output pair and produces a probability of the how realistic they look. The weights of the D, are then adjusted based on the classification error of the input output pair and the input target pair. G weights are then adjusted based in output of D. As what as the difference between the output and target image, both D and G will improve overtime during this process.
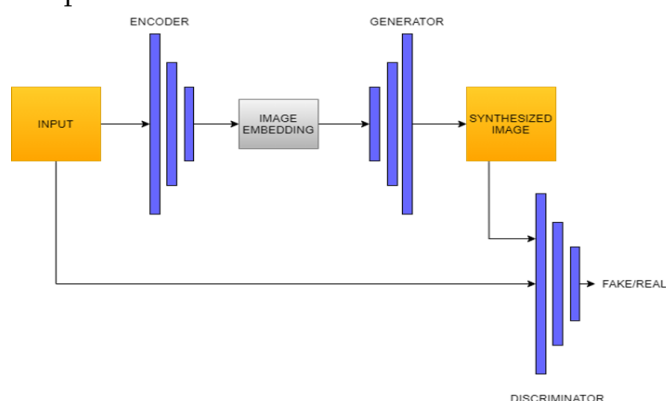
Figure 3. GAN

C. Results

The datasets that are used in this method firstly we need to extract the skeleton figure of the frames obtained from the video. The skeleton is obtained by using pose estimation method that is open-pose.

Figure 4. Pose Estimation

The dataset is then sent to the generator where the target image and the pose stick figure is synthesized and sent to discriminator. This process is cycled until the final result is obtained.
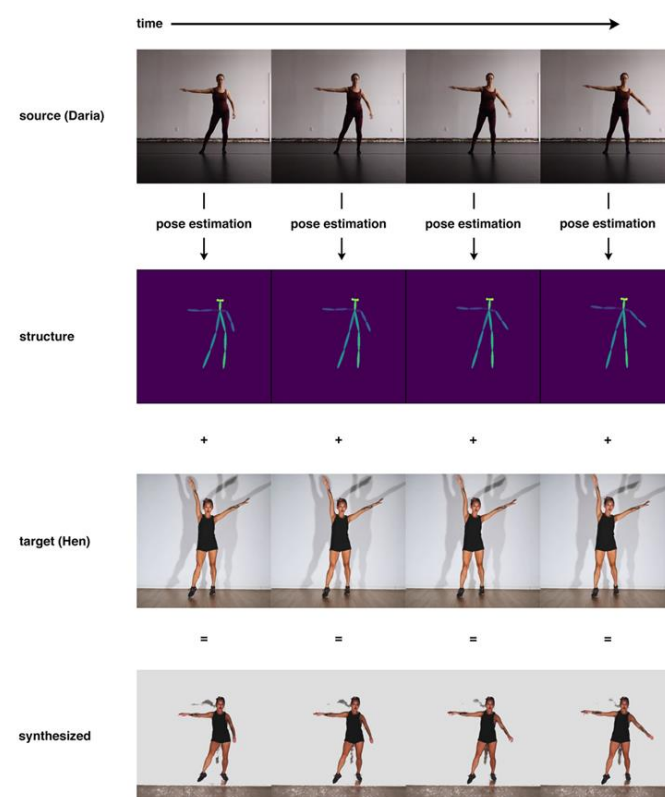
Figure 5. Expected Result

VII.   SCOPE OF THE PROJECT

Human-pose-estimation is one of the key problems in the world of computer vision that has been studied well over the past 15 or for more years. The reason for it's important is the abundance of applications that can be benefitted from such a technology.

The scope of future research in this project is immense and will create a learning slope that can get more people interested. The detecting and tracking of pixels basically representing humans is known as human motion capture.

Human motion captures the digitized human motion, allowing the machines to track or reconstruct the human behaviour. The main advantage of this is that the large amounts of human motion data can be processed within few milliseconds. This enables an application that can be performed in a real time, such as movement analysis for sports. It is vastly used in health research and kinesiology to help people improve their posture, gait and other movements.

Motion capture is performed via joint skeletal tracking which tracks the humans in a video or in an image by creating a virtual joint and connecting the joints to get a skeleton. Together, these joint skeletal data allow us to analyse the pose and the movements and then reconstruct the human behaviours.

The main reason for this trend is the ever-increasing applications. That could be human-robot interaction, gaming, sports performance analysis etc.

Nowadays deep fake has been one of the biggest issues of the present generation. Deep fake is nothing but the manipulation of the videos or images or the other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds that appear to be real.

VIII.   PROGRESS AND LIMITATIONS

As this technology grows in future, this could be used as fun time as snapchat filter but as it can better, it will be useful for movie studios, music videos, advertising campaigns using it to help enhance their production in a low-cost way. Why hire entire background dance crew when you can just generate them. Also lead actor won't need any intense training to record their dancing sequence and because it can lower cost so drastically expect to see better quality content coming from the amateur video production community as well in a way it democratises dance.

The dark side of this is that just like deep fake technology is moving towards so that many videos can be manipulated. People can be framed for crimes that they didn't commit. Can use AI to synthesis voices, faces, photos and now video.

So, we need to create new methods of identity verification since video whether from smart phone or cam is the key to make or break of many defence cases.

One solution is to use AI to fight AI. Another solution is blockchain. A new start-up called factom has a solution that's already being tested by the department of homeland security. Since the bitcoin blockchain is an immutable data Structure so that no one can modify. Factom is using it to time stamp video of data from say a security camera at specific intervals. Because it digitally signs and hashes data instantly. The pixels are pull of camera they can confidently claim that are video was really taken by camera that digitally signed the data. In general, we can all use the blockchain to digitally sign and confirm the authenticity of the video file that relates to us. The more digitally sign added more valid it is considered.

IX. CONCLUSION

Human pose estimation has been evolving discipline with opportunity for research across various fields. Due to their superior performance across the tasks and data set they have been in a great demand to be applied in the fields of machine learning and deep learning. Several systems based on the presented prototype description were lately successfully employed commercially. One example is an

augmented reality stage performance art piece where a 3D rendered dancing figure like person appears to float next to a real dancer. Another is an in-game entertainment application making NBA players dance. Some losses are found in the final result which could be possibly resolved by using extra GAN and temporal smoothing.

X. ACKNOWLEDGEMENT

XI. REFERENCES

[1]. M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[2]. Michael Gleicher. Retargetting motion to new characters. In Proceedings of the 25th annual conference on Computer graphics and interactive techniques, pages 33–42. ACM, 1998.

[3]. Dan Casas, Marco Volino, John Collomosse, and Adrian Hilton. 4D Video Textures for Interactive Character Appearance. Computer Graphics Forum (Proceedings of EUROGRAPHICS), 33(2):371–380, 2014.

[4]. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In CVPR, 2017.

[5]. Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pages 353–360. ACM Press/Addison-Wesley Publishing Co., 1997.

[6]. Umar Iqbal, Pavlo Molchanov, Jan Kautz, "Weakly-Supervised 3D Human Pose Learning via Multi-View Images in the Wild", Computer Vision and Pattern Recognition (CVPR) 2020 IEEE/CVF Conference on, pp. 5242-5251, 2020..

[7]. Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexie A. Efros. Everybody Dance now. In IEEE Conference on Computer Vision, 2019.

[8]. Jogendra Nath Kundu, Siddharth Seth, Varun Jampani, Mugalodi Rakesh, R. Venkatesh Babu, Anirban Chakraborty, "Self-Supervised 3D Human Pose Estimation via Part Guided Novel Image Synthesis", Computer Vision and Pattern Recognition (CVPR) 2020 IEEE/CVF Conference on, pp. 6151-6161, 2020.

[9]. Darioo Pavllo, Christoph Feichtenhofer, David Grangier, Michael Auli, "3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training", Computer Vision and Pattern Recognition (CVPR) 2019 IEEE/CVF Conference on, pp. 7745-7754, 2019.

Classification of Soil Contamination using Machine Learning

Shivani Patil[1], Kunal Sarode[1], Rakesh Suryawanshi[2]

[1]Department of B E CSE, A C Patil College of Engineering, Maharashtra, India

[2]Professor & Head, Department of B E CSE, A C Patil College of Engineering, Maharashtra, India

ABSTRACT

Agriculture is the tree whose roots are soil, that provides food to all the living organisms present on this planet Earth. It is indirectly responsible for existence of humans on this planet. As we have learnt in our primary as well as secondary schooling that soil is a non-renewable thing, so here comes the question that how are we going to conserve it?? Looking at the results and graphs on soil, it is evident that soil contamination has crossed its limits and there are major needs to minimize the soil contamination. Presently, there are numerous ways to reduce soil contamination but none of them are optimized. So, there should some solutions that is feasible to farmer as well so that the quality and quantity of the crop does not get affected. In the Computer Science world of Technology, we are having a topic, namely, Machine Learning, that can play a major role in finding the optimized solution for soil contamination. Machine Learning is the upcoming technology, and experimenting some ideas in that domain is the interesting thing. So, we have clubbed these two topics and made a final year project titled "Classification of Soil Contamination using Machine Learning". In our project, we are going to predict whether soil entries that we have considered is naturally or man-madly contaminated considering the condition, metal proportions. More specifically, the process includes three main modules: Data Pre-processing, Training and Testing Modules. The secondary objectives that we are going to include are crop prediction and suitable fertilizer is being suggested.

Keywords—Metal Proportion, Imputer, Contamination, SVM, Decision Tree, RF, Gaussian Naive Bayes, Logistics Re- gression, Crop Prediction, Fertilizer

I. INTRODUCTION

Every living organism on this planet, in all sense, depends on soil. All the necessary components required for the living organism such as food, water, air is somewhere connected to the soil. The food full of nutrients are healthier than that of lack of nutrients. The whole food chain has a primary component, i.e., Soil. But do we take care of that resource which provides us food?? The answer is no because it is getting contaminated in numerous ways. In simple words, Soil contamination can be stated as the disturbances in the soil due to absence or presence of physical, chemical and biological factors. There are numerous ways present to treat the soil, making it fertile. But those ways are meaningful only when we get the exact cause of the contamination. Here our project gets introduced wherein we are going to

predict whether a given soil sample is naturally contaminated or unnaturally contaminated on the basis of proportions of metals, also crop recommendation and fertilizer suggestion are the secondary functions included.

II. LITERATURE REVIEW

There were some minor projects that were carried out on the topic related to our project. All the studies were having some major demerits making those project to give less fruitful results. Some of them are stated below.

A paper titled "Soil Quality Prediction Using Machine Leaning Techniques" was approached by T. Venkat Narayana Rao. It consisted the prediction of soil quality using two supervised models, i.e., Random Forest and Decision Tree Algorithm considering all three parameters (physical, chemical, biological). The result proclaimed that the prediction was less optimal, the reason being, more parameters, less data-set, improper selection of algorithms.[1]

Another approach was by Faridon Ghadimi, titled "Prediction of Heavy Metals: A Case Study in Iran". The prediction was done considering only heavy metals, except macro-nutrients, using two supervised algorithms namely, SVM and lastly, k-NN. The result proclaimed that the desired outcome was fair enough but not satisfactorily due to consideration of less attributes.[2]

One more approach was proposed by Aleo Mok, titled "Classification of Soil Contamination". The ideology of this project was to predict whether a soil sample, of McConnell Air Force Base, is naturally contaminated or unnaturally on the basis of metal proportions. The result proclaimed that he got prediction with some 85 accuracy with a conclusion that soil is unnaturally contaminated. The major demerit of this project was less data-set as he predicted only for one soil sample that significantly reduced the optimality of the project.[3]

III. THE DATA MINING PROCEDURE

The basic architecture of the machine learning is as follows:
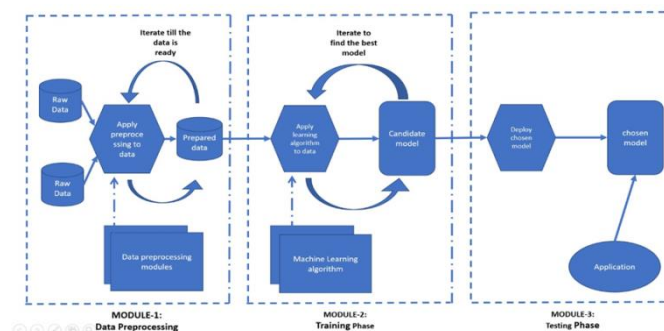
Figure 1. Machine Learning Architecture

The skeleton of machine learning architecture is stated below: Figure 1 shows the procedural architecture of machine learning. The methodology of the proposed system comprises of the following steps:

- Data-set Collection.
- Pre-Processing.
- Training and Validating.
- Prediction.
- Result.

Let's see some brief description on the above-mentioned topics:

A. Data-set collection
The most root part of machine learning process is data-set. So, here we have a valid data-set of Karnataka State having 92,832 soil samples of various villages. Our soil data-set consists of 92,832 rows and 17 columns and our crop data-set consist of 2200 rows and 8 columns.

B. Pre-processing
In machine learning, we use "pandas", "NumPy", "mat-plotlib", "seaborn" libraries for pre-processing purpose.
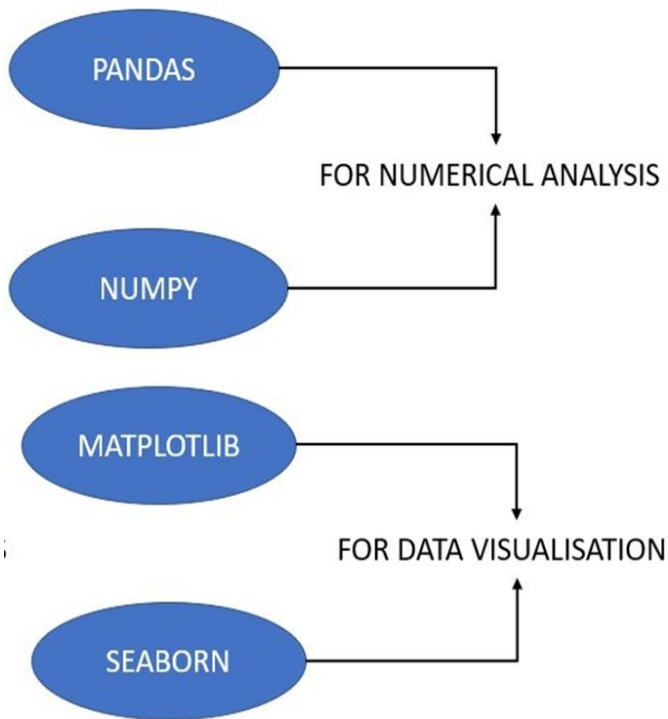
Figure 2. Machine Learning Architecture

NUMPY: NumPy, 'Numerical Python' or 'Numeric Python', is machine learning module used to the total size of data-set. It helps user to do fast mathematical computation on arrays and matrices. Syntax of Numpy is as follows:

>>>import numpy as np

PANDAS: Pandas is the most powerful module or library in machine learning that helps us to give a brief idea about the data-set used. It provides user friendly tools reducing the complexity and dimensionality of codes in machine learning. Pandas is an excellent feature through experimenting different ideas on machine learning becomes easy. Syntax of Pandas is as follows:

>>>import pandas as pd

MATPLOTLIB: Matplotlib is an open-source data visualization technique used for plotting graphs, correlation matrices, etc. It holds the feature for visualizing the huge data-sets and that becomes a helping hand for user to choose the appropriate models and get the accuracy required for the project. Matplotlib can be imported into Python using:

>>>import matplotlib.pyplot as plt

Through matplotlib, we have created histograms, scatter plots, correlation matrix, grid correlation matrix, etc.
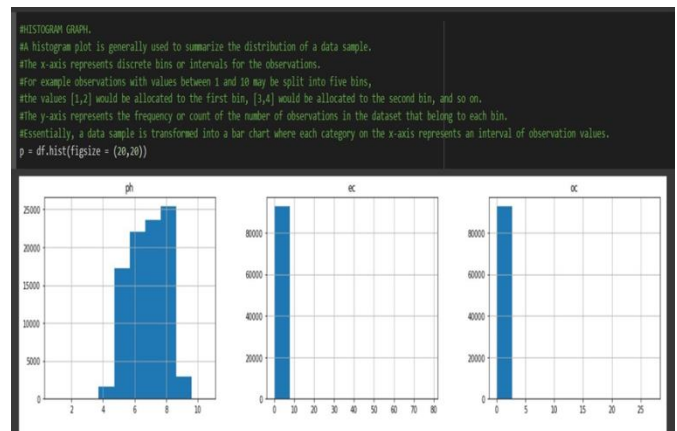
Figure 3. Histogram representing the frequencies of various metal proportions of the given soil data-set

SEABORN: Seaborn is a version of matplotlib. It's a data visualization technique that gives you a high-level GUI to create visually appealing and insightful statistical graphics. Seaborn can be imported into Python using:

>>>import seaborn as sns

Seaborn can be implemented for: Visualizing statistical relationships and Plotting categorical data. Seaborn can be used to build graphs, correlation matrix, grid correlation

, matrix. denoting the missing entries by "yellow" color.
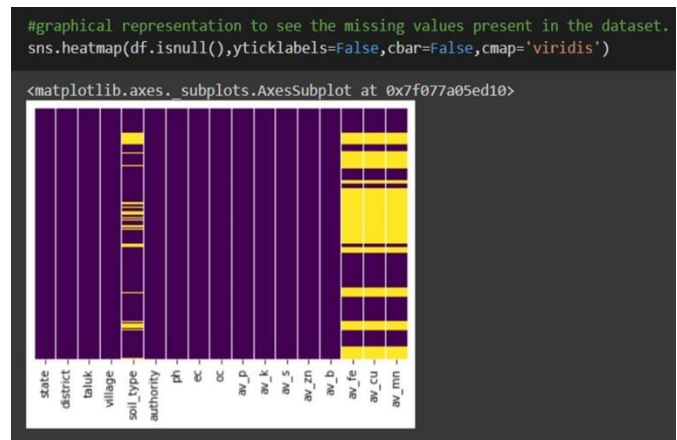
Figure 4. The above diagram gives us brief knowledge about data entries, denoting the missing entries by "yellow" color.

C. Training and Validation

Now, the processed data are stored in ".csv " file for further use. The pre-processed data-set of the Karnataka soil sample is divided into two parts :

- Training.(70 % of the data-set is used)
- Testing.(30 % of the data-set is used)
- Syntax

```
# Splitting into train and test data

from sklearn.model_selection import train_test_split
Xtrain, Xtest, Ytrain, Ytest = train_test_split(features,target,test_size = 0.3,random_state =2)
```

Figure 5. Training and Testing of data-set.

Now, comes the training part of the models. So, classification models are trained and tested to get the accuracy of the models. Once, done with the accuracy part, we need to perform cross validation for further efficiency of the project.

D. Prediction:

Based on different supervised models, we got some ac- curacy's and predictions that helped us to reach the desired outcomes.

```
#Making a prediction

data = np.array([[7.4,  0.10,  0.18,  3.6,  53.0,  2.6,  0.24,  0.18,  27.463554, 1.938304,  16.995547]])
prediction = RF.predict(data)
print(prediction)

['unnatural']

#Making a prediction

data = np.array([[7.4,  0.31,  1.06,  70.3,  269.0, 19.7,  4.96,  0.70,  17.791627, 1.677842,  13.865944]])
prediction = RF.predict(data)
print(prediction)

['natural']
```

Figure 6. User prediction.

E. Result:

The final result gives the prediction of soil

IV. CLASSIFICATION ALGORITHMS

Some of the classification algorithms used in soil con- tamination and crop recommendation are SVM, LOGISTIC REGRESSION, RANDOM FOREST, GAUSSIAN NAIVE BAYES, XGBOOST CLASSIFIER and DECISION TREE.

A. SVM:

Support vector machine is the most common classification algorithm that falls under the category of supervised machine learning model. Since, our project comes under the classification problems, we are using this model for better results. Each entry in the data-set is considered as a unique observation along with some coordinates, this all sums up and referred as" Support Vectors". SVM involves a process of dividing the two classes with a hyper-plane. The only aim that user have is to find the exact and the right hyper-plane. The discern under suggests the choice characteristic for a linearly separable hassle, guide vectors:
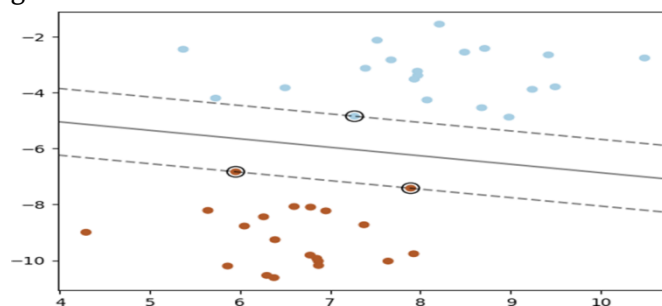
Figure 7. Graphical representation of SVM

In SVM, we are using SVC type,

Given training vectors $x_i \in \mathbb{R}^p$, in two classes, and a vector $y \in \{1, -1\}^n$, our goal is to find $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ such that the prediction given by $\text{sign}(w^T \phi(x) + b)$ is correct for most samples.

SVC solves the following primal problem:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$

$$\text{subject to} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$

$$\zeta_i \geq 0, i = 1,\ldots,n$$

Intuitively, we're trying to maximize the margin (by minimizing $||w||^2 = w^T w$), while incurring a penalty when a sample is misclassified or within the margin boundary.

B. LOGISTIC REGRESSION:

Logistic Regression is the type of supervised machine learning model. It is the model in which the prediction output has only two values, like in our case, we have two values (naturally and man-madly contaminated.). The data-set on which the prediction is to be made has features (independent variables) and target (dependent variable). Logistic regression is

subset of linear regression where we predict whether the soil is naturally contaminated or man-madly and also crop recommendation is done. The mathematical part of this model contains the log function.

Logistic Regression equation:

$$p = 1 / 1 + e^{-(\beta 0 + \beta 1X1 + \beta 2X2 .... + \beta nXn)}$$

C. RANDOM FOREST ALGORITHM:

Random Forest is a, supervised machine learning model, classifier. It is a collection of numerous decision trees on different sub-sets and take the average of all the decision tree. This complete procedure is based on the data-set that we have selected for this project. The main motive of this project is to improve the accuracy of the data-set. Random forest algorithm is most advantageous as it has a main feature of considering the prediction result from all the trees with a concept of majority votes, and providing us the desired outcome. Because of that feature, if there are a greater number of trees that will give us perfect accuracy, eliminating the problem of over fitting.

The Mathematics that makes Random Forest algorithm work

Regression Model

The mean squared error (MSE) is used to branch your data from each node.

$$MSE = \frac{1}{N} \sum_{i=1}^{N}(fi - yi)^2$$

Where N is the number of data points, fi is the value returned by the model and yi is the actual value for data point i.

Classification Problems

While using Classification Problems in Random Forest we use Gini Index. The workflow of this formula considers the class and probability to determine the Gini of each branch on a node having a count of which branch is most occurring one. To get some more brief idea about prediction in this section, we can include entropy to see how branching of nodes is done in decision tree.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

To check which branch is more likely to occur, we use Gini of all branch on different node, considering class and probability. Another mathematical method present in decision tree is entropy.

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2 (p_i)$$

D. DECISION TREE ALGORITHM:

Decision Tree plays a very important role as a supervised model, because it consists of internal nodes denoting features of our data-set, branches denoting the decision rules and output is represented in the form of different leaf node. The second name of decision tree is referred as tree-structured classifier. This algorithm aims to get all the possible solutions considering the given conditions in a graphical way. The major disadvantage of decision tree is the selection of the best attribute for all the types of nodes, including the root node and the sub-nodes. To solve the major disadvantage of decision tree, we have a solution or a method referred as" Attribute Selection Measure (ASM)". This method helps us to select the best attribute required for the nodes of the tree. The two main strong pillars of Attribute Selection Measure are listed below:

Information Gain= Entropy(S)- [(Weighted Avg)

*Entropy(each feature)

Gini Index

$$Gini\ Index = 1 - \sum_j P_j^2$$

E. XGBOOST CLASSIFIER:

XGBoost algorithms is the most promising supervised ma- chine learning algorithm of the type" Gradient Boosted Trees algorithm". It works on the principle of

function approximation with two in-built features: optimization of specific loss functions and numerous regularization techniques.

The objective function that aims for reducing the purpose is stated below:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

Real value (label) known from the learning data set ↑

Can be seen as f(x + Δx) where x = $\hat{y}_i^{(t-1)}$

XGBoost objective function analysis

F. GAUSSIAN NAIVE BAYES:

Naive Bayes is supervised machine learning algorithm. Naive Bayes, being a classification algorithm, is used when the target variable (output variable) is discrete. The Naive Bayes Algorithm is implemented in two parts:

The Naive Bayes formula (Theorem): The Naive Bayes Theory in the most cases can be reduced to a formula:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

A Distribution (in this case Gaussian one): This part of the algorithm works on two elements (the mean and the standard deviation). These two elements help us to fins its distribution function.

It has the next form:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

G. CROSS VALIDATION:

Cross-validation is a module in which we divide the remaining data into different parts and then we try to find the accuracy of that different parts. If all the accuracies are same, it means we have successfully completed the project with right conditions and assumptions. Cross Validation needs to be executed in following order:

Step1: Divide the data in 7: 3 ratios.

Step2: Use maximum part for training

Step3: Use the remaining part for validating to see whether the requirements of the project are satisfied or not.

V. PROPOSED METHODOLOGY
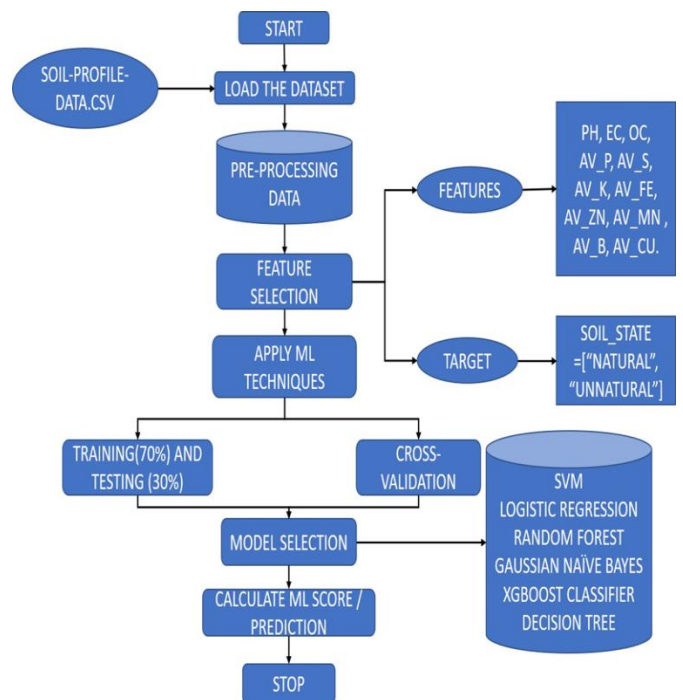
Figure 8. Architecture of Proposed System of soil contamination

The method of the challenge is designed in six steps:

- Installing the Python and SciPy platform. We want to mount our ".ipynb" report on our google force for simi- larly access.
- Importing the data-set. The data-set of Karnataka soil is needed to be imported in ".csv" format.
- Summarizing the data-set. Sorting and cleaning of data is the mandatory process to increase the efficiency of the project. We can fill the missing data using "imputer" function.
- Visualizing the data-set. Analyzing 92,833 data-set man- ually is quite difficult task, so we can

visualize it through different graphs such as histogram, scatter-plots, log graph, etc.

- Evaluating the algorithms. After visualizing the data-set, now comes training and testing part!!! Let's divide the data into 7:3 ratio where 70% of the data will be trained and 30% will be tested. Now, let's select the appropriate models and then train them to get the accuracy of the prediction. We have used 6 models: SVM, Logistic Regression, Random Forest, Decision Tree, XGBoost, Gaussian Naive Bayes. After getting the accuracy of each model and comparing them, lets cross validate to see the efficiency of the models.

- Making some predictions. Now, comes the last stage of the project, i.e., to make predictions. Here, user can man- ually give the input and get the prediction whether soil is naturally contaminated or unnaturally contaminated.
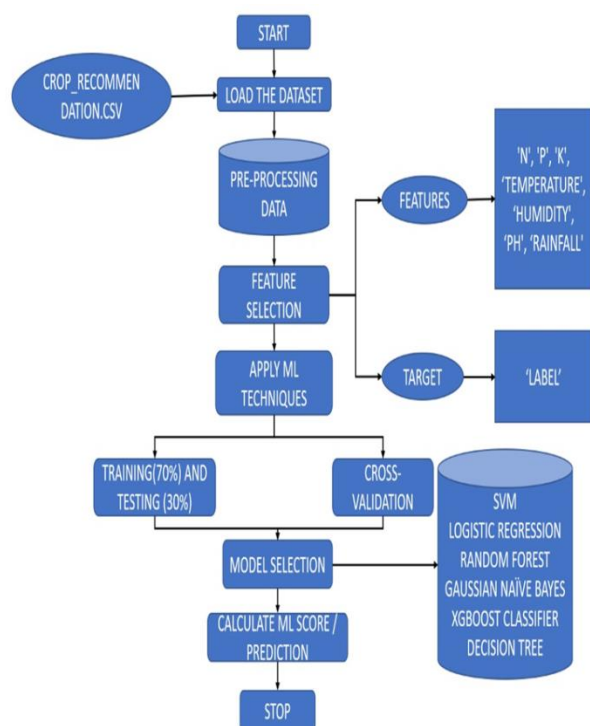
Figure 9. Architecture of Proposed System of crop recommendation

The method of the challenge is designed in six steps:
- Installing the Python and SciPy platform. We want to

- mount our ".ipynb" report on our google force for similarly access.

- Loading the data-set. The data-set of crop recommendation is needed to be imported in ".csv" format.

- Summarizing the data-set. Sorting and cleaning of data is the mandatory process to increase the efficiency of the project. We can fill the missing data using "imputer" function.

- Visualizing the data-set. We can visualize our" crop recommendation.csv" data-set through different graphs such as histogram, scatter-plots, log-graph, etc.

- Evaluating a few algorithms. After visualizing the data- set, now comes training and testing part!!! Let's divide the data into 7:3 ratio where 70% of the data will be trained and 30% will be tested. Now, let's select the appropriate models and then train them to get the accuracy of the prediction. We have used 6 models: SVM, Logistic Regression, Random Forest, Decision Tree, XGBoost, Gaussian Naive Bayes. After getting the accuracy of each model and comparing them, lets cross validate to see the efficiency of the models.

- Making some predictions. Now, comes the last stage of the project, i.e., to make predictions. Here, user can manually give the input and get the prediction of crop.

- Now, to make the project more user-friendly, we have designed a front-end as well!!The front end consists of a website with three functions, namely, contamination, crop and fertilizer. The front end can be created using flask module in python and few HTML files to link.

VI. RESULT ANALYSIS

The building blocks of the proposed model are soil data-set and crop data-set. For best accuracy's, we have implemented the project using most suitable supervised machine learning models considering the metal proportions present in the data- set. The

expected results were: To predict whether the soil is naturally contaminated or man-madly contaminated and to predict which crop is suitable. From the experimental result, we see that XGBoost classifier has obtained the maximum ac- curacy for soil contamination and Decision Tree has obtained maximum accuracy for Crop Prediction.

The classification accuracy is tabled below:

Sr No	Dataset	Algorithm	Accuracy
1	92831	SVM	96%
2	92831	Logistic Regression	97%
3	92831	Decision Tree	98%
4	92831	Gaussian Naïve Bayes	86%
5	92831	Random Forest	99%
6	92831	XGBoost Classifier	100%

Figure 7. Accuracy chart of soil contamination

Sr No	Dataset	Algorithm	Accuracy
1	92831	SVM	11%
2	92831	Logistic Regression	95%
3	92831	Decision Tree	90%
4	92831	Gaussian Naïve Bayes	99%
5	92831	Random Forest	99%
6	92831	XGBoost Classifier	99%

Figure 8. Accuracy chart of crop prediction

VII.   CONCLUSION AND FUTURE ENHANCEMENT

Soil pollutants is the end result of quite a few events, inclusive of herbal screw ups and human experimentation, each of which land up contaminating the soil. Excess content material of heavy metals or extra content material of micro nutrients, fertilizers and pesticides are the maximum not unusual place reasons of synthetic soil infection. From the analyses of the data with respect to classification, it can be stated with high confidence that the Karnataka soil has natural soil contamination. This contamination is characterized by especially normal concentrations of micro nutrients. Also, we have successfully implemented crop recommendation part as well!! In the future scope, we are expecting to

predict the season as well, with the crop prediction, we are also thinking of evaluating some other algorithms to get more prediction Also, we are expecting to add one more feature that is disease prediction of crop.

VIII.   ACKNOWLEDGEMENT

IX. REFERENCES

[1]. T Venkat Narayana Rao and S Manasa. Artificial neural networks for soil quality and crop yield prediction using machine learning. Iranian Journal of Energy and Environment, 2019.

[2]. Faridon Ghadimi. Prediction of degree of soil contamination based on support vector machine and k-nearest neighbor methods: A case study in arak, iran. Iranian Journal of Energy and Environment, 5(4):0–0, 2014.

[3]. Aleo mok, Classification of Soil Contamination, 2014.

Searching for an Exoplanets Using AI

Shruti Akashe[1], Pranita Banda[1], Suyog Dale[1], Mrs. Shilpali Bansu[2]

[1]B.E. Student, Department of Computer Science and Engineering, A C Patil College of Engineering, Maharashtra, India

[2]Assistant Professor, Department of Computer Science and Engineering, A C Patil College of Engineering, Maharashtra, India

ABSTRACT

Various scientists have used NASA's Kepler Space Telescope for several years to discover thousands of new stars and exoplanets. During the Extended K2 mission, they detected various types of stars and exoplanets in various regions of the sky and, as a result, in various galaxy environments. Astronauts are interested in learning more about the population of various Exoplanets in different systems. Whatever they need is an automated and unconditional method of detecting an exoplanet in the vicinity and producing a false positive signal that transmits the planet signal. We have a method for identifying an exoplanet that uses deep learning, a type of machine learning algorithm that has become well-known in the field of linguistics. We previously used neural networks to classify an exoplanet in the K2 region, and we will continue to use neural networks, as well as other algorithms such as data augmentation and artificial neural networks, to improve accuracy. That means better precision and recall values for exoplanet discovery. K2 excels at this mission, achieving a 98 percent accuracy rate. As a result, although it is effective at detecting False Positives, it still requires human supervision to generate a full star sample. We previously used a variety of algorithms and Kepler data, but now we will use a cutting-edge technology algorithm called a Neural Network to classify new stars and exoplanets, as well as their population dependence.

Keyword — Deep Learning, Exoplanets, prediction, Star Brightness, Feature Engineering, Data augmentation, Neural Network, 1D Convolutional, ANN, RNN.

I. INTRODUCTION

Conveyancing Exoplanets do provide a remarkable Oppor-tunity to ascertain planetary atmosphere through the various spectroscopic features. During Various primary transit, when a planet goes nearby or passes nearby it's host star, the light which transmits the planet's atmosphere that reveals absorption from the atomic and molecular species. So for that, with the recent activity in technology, we can also use the famous technology which is powerful telescope, i.e NASA's Kepler Space Telescope, for observing the stars and planets which is visible through universe for the extrasolar stars and extrasolar planets which are also called as Exoplanets.

II. SCOPE

While doing this work, we address many of the task which creates an automatic system which identifies

the planet, researcher in Kepler's K2 data with the help of deep learning, which is a modern machine learning technique. Recently we have seen different machine learning algorithms which is used for other algorithm related purposes. The consisting machine learning algorithm for the classification of planet Researcher and False positives result in the Kepler's data which has to Autovector Project at a random forest classifier which classifies algorithm made decision concluded by metrics which is obtained by the K-Pipeline.

III. PROBLEM STATEMENT

Few decades ago, over a million stars which were monitored to detect the conveying planet's. The One-by-one interpretation for an potential exoplanet candidate in which there is fine intensive and Subject to human being error, this types of error are so difficult to clarify and solve. So here we present a best way to detect an exoplanet in large number of planetary project, unlike the current methods used the best algorithm neural network.

IV. LITERATURE REVIEW

Paper 1:
A model trained on deep neural networks is presented in the paper "BAYESIAN DEEP LEARNING FOR EXOPLANET ATMOSPHERIC RETRIEVAL," by Frank Soboczenski and Michael D. Himes. They used convolutional neural networks in one dimension. To generate predictive distribution, they also used Monte Carlo dropout approximation. While the approach yielded positive results, the quest for the best model remains unfinished.

Paper 2:
Liang Yu and Andrew Vandenberg's paper "IDENTIFY-ING EXOPLANETS WITH DEEP LEARNING III. AU-TOMATED TRIAGE AND VETTING OF TESS CANDI-DATES" describes a deep learning model capable of conduct-ing TESS candidate triage and vetting. It is the first neural network to be trained and tested on real TESS data, and it is based on an existing neural network. With an average precision of 97 percent and accuracy of 97.4 percent, the design can discern transit-like signals from stellar variability and instrumental noise.

Paper 3:
Christopher J. Shallue and Andrew Vanderburg's paper "IDENTIFYING EXOPLANETS WITH DEEP LEARNING: A FIVE PLANET RESONANT CHAIN AROUND KEPLER-80 AND AN EIGHTH PLANET AROUND KEPLER-90" introduces a five-planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. TensorFlow, an open source machine learning software library, was used in the model. To reduce the cross-entropy error function, they used the Adam optimization algorithm. Dropout regularization was also extended to the completely connected layers. For black-box optimization, they used the Google-Vizier method. Two of the new candidates, Kepler-80 and Kepler-90, were statistically validated by the model.

Paper 4:
The final K2 model – Astronet K2 is a one-dimensional convolution neural network with max pooling – is presented in the paper "IDENTIFYING EXOPLANETS WITH DEEP LEARNING II: TWO NEW SUPER-EARTHS UNCOVERED BY A NEURAL NETWORK IN K2 DATA" by Anne Dattilo and Andrew Vanderburg. To reduce the cross entropy error function, the model used the Adam optimization algorithm. On the research data collection, the model had 98 percent accuracy and an AUC of 0.988.

V. PROBLEM DEFINATION

A. Objective

We've come to show you a new approach for detecting exoplanets using neural networks. Deep learning is another term for neural networks. Deep learning is a technique for creating a machine view of a problem by teaching the algorithm with a variety of models.

B. Overview of Proposed System

1) We will be using Kepler dataset for analysis.
2) Then we will do feature engineering, and extraction of features.
3) Then we will do baseline model training then after evaluation of the model and assessing the performance.
4) We will do feature scaling and remove some unwanted features.
5) Then we will again train our model. We will be using neural networks for our expected results.

VI. PROPOSED SYSTEM

A. Architecture

Figure 1. Architecture Of the Proposed System.

So, in architecture we are providing a raw data set to the model then the following operations are going to be performed on that data.

1) Data Partitioning: In Data Partitioning, the data will be divide into the testing data, validation data, and training data. then that datasets are passes to the data IO.
2) Data IO: It contains the input output function which has to performed on that data.
3) Data Sampling: In data sampling testing dataset will be passed to the evaluation and validation data and training data will be further proceed to data augmentation.
4) Data Augmentation: Data augmentation, is used to increase the size of dataset. So, further validation samples and training samples are passed to the model selection and model fitting. In that for loss , we will be using binary cross entropy function .and for optimization we will be using Adam optimizer. Then we will be taking trained models from model zoo and trained parameters along with a model and hyper parameters for comparison. then after comparison we will get results in evaluation ,then it will further pass to model zoo for final validation.

Figure 2. Data Augmentation.

B. Dataset

1) Training Testing dataset:

- Training dataset: 5087 rows (observations) and 3198 columns (features), with column 1 being the mark vector. Columns 1 and 2 show the shift in flux values over time.
- There will be 37 confirmed exoplanets and 5050 unconfirmed exoplanets in the training dataset.
- Testing Dataset: 570 rows (observations) and 3198 columns (features), with the mark vector in column 1. Columns 1 and 2 show the shift in flux values over time.

- There will be 5 confirmed exoplanets and 565 unconfirmed exoplanets in the testing dataset.
- Either star has a binary value of 2 or 1, with 2 indicating that the star has a known exoplanet in its orbit and 1 indicating that the star does not.
- The luminous flux of a star is used to calculate its brightness.
- The light curve intensity observed for each star at various points in time is luminous flux 1-flux 3197.

C. Feature Extraction:

The function extraction can be done in two ways. The first uses manual techniques for feature creation, while the second employs automated techniques for feature generation. When looking for a light curve, both approaches are used.

1) Manual feature extraction: We used extraction tech-niques that are specialized on time series data, and we used a python library called Feature Analysis for Time Series to extract manually (FATS).

2) Automatic feature extraction: We use unsupervised learning methods in automated feature extraction techniques to find intrinsic patterns among all the data, which is independent of the task.

D. Neural Network:

Artificial neural networks are the functional unit of deep learning and are one of the most important topics in artificial intelligence. Deep learning assists AI by supplying a collection of algorithms and neural networks for problem solving. When any input data is provided to a neural network, the data is processed through layers of perceptron to produce the desired output.

Starting from the left, we have:
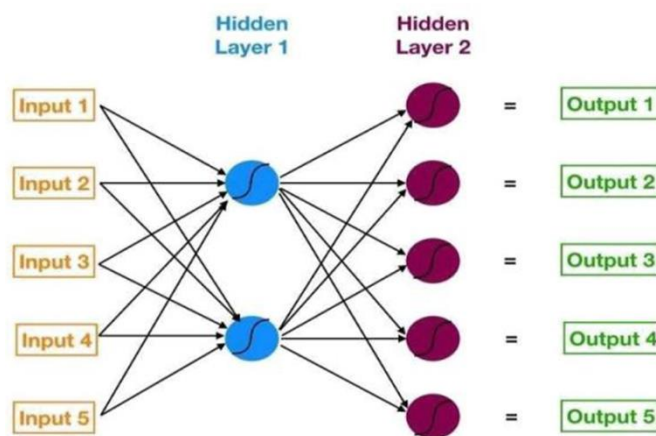
Figure 3. Neural Network:

E. Metrics:

In performance measure metrics ,accuracy is not enough to evaluate the algorithms.so that's why we are evaluating the model with the help of other metrics like Recall,Precision,F1 score.

Precision : Precision is the ratio of true positives in our tests to all positives in the samples (true positives or false positives).

$$P = \frac{T_p}{T_p + F_p}$$

Recall: The ratio of true positive values in our samples to the total samples is called recall. As a result, a higher recall value indicates that we received more positive samples.

$$R = \frac{T_p}{T_p + F_n}$$

F1 score: The F1 Score is calculated by adding the recall and precision values together. F1 Score has a best value of 1 and a worst value of 0.

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$

F. Models And strategies:

1) 1D Convolutional model: In ANN model evaluation we are using 1D Convolutional model .A modified form of 2D CNN's is called as 1D Convolutional neural network. 1D CNNs are having more advantages than 2D that's why we are using 1D CNNs over here. In that we are passing the fitting model ,testing and training parameters. True positive, true negative, false positive, and false negative data values are used in the Confusion Matrix. We use convo1D which means an operation which is used to summarize a tensor values .we use relu activation function which is the equation to determine the output of a neural network ,and Relu means a Rectified Linear Unit which provides us an output if the star is having an exoplanet. also we are defining the Maxpooling1D,which take the maximum value defined by the pool-size. for the compilation of the model we use Adam optimizer that handle the noisy problems or the noise in our model and We use binary cross entropy for the loss, which is the loss calculated for the CNN output component and is unaffected by the values of the other components.
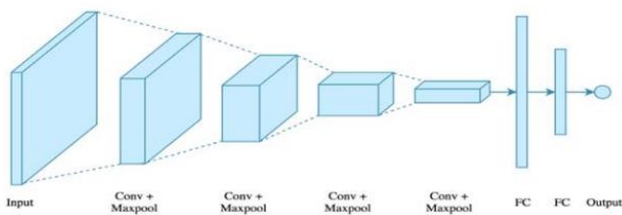
Figure 4. Convolutional Model.

2) Data Augmentation: In Data Augmentation ,we can increase the size of dataset .We are creating a new training Dataset with the help of existing Training dataset. before doing data augmentation the size of dataset is (3560,3197),and after doing data augmentation over the dataset the size of new dataset is (6460,3197).

3) Scaling: The reason behind the scaling is as you know different stars have different starlight(star brightness), so it will better to scale each star data that's why we use scaling over here.

4) RNN Model:: The Recurrent Neural Network (RNN) is a form of neural network that can be used to predict future events. A Recurrent Neural Network is a form of neural network that saves the output of one layer and feeds it back into the input to produce an output layer.
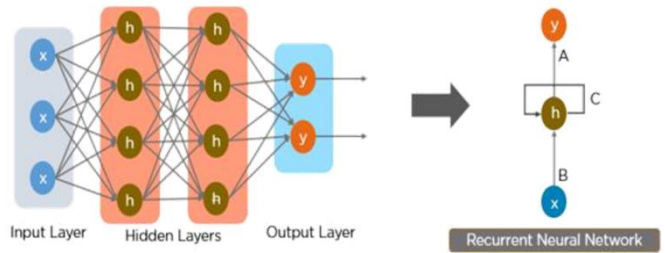
Figure 5. RNN Model

G. Analysis And design Consideration:

Use Case Diagram:
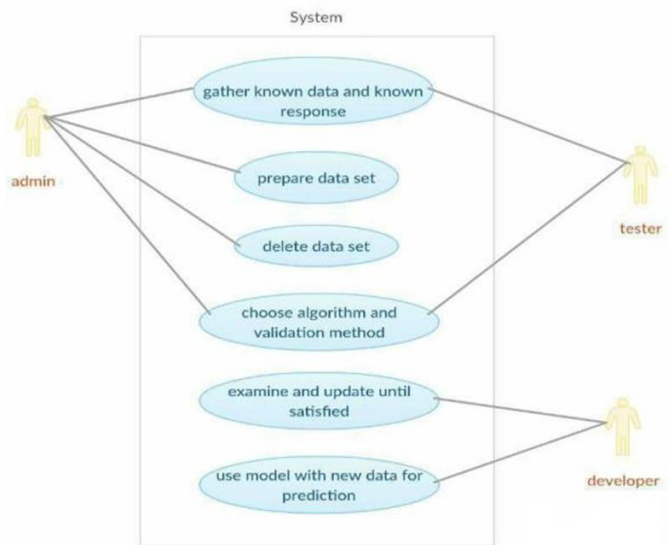
Figure 6. Use Case Diagram

Activity Diagram:
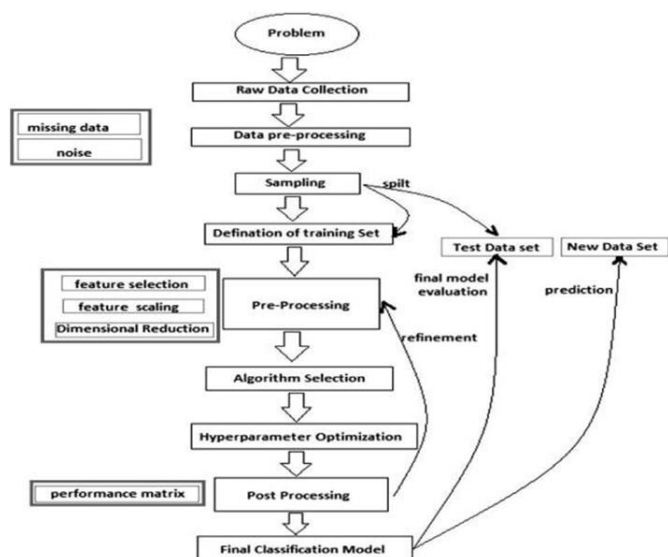
Figure 7. Activity Diagram

H. System Requirements:

1) Software Requirements:
- Google Colab
- Language: Python
- Libraries: matplotlib, pandas, scikit learn, TensorFlow, NumPy
- Architecture: neural networks

2) Hardware Requirements:
- Nvidia Geforce GTX graph-ics min 6 gb
- Ram 16 gb
- Processor intel core i7 10th Generation

VII. RESULT ANALYSIS

The proposed model results is based on the flux values. we used several deep learning models to detect whether the star is having exoplanet or not. The following tables shows a final conclusion of our model.

Model	Test	Train
1D Convolutional model	99% accuracy with no precision and recall	99% accuracy with no precision and recall
Data Augmentation	99% accuracy with no precision and recall	97% accuracy with precision 100% and recall 94%
Recurrent Neural Network (RNN)	89% accuracy with recall 38%	66% accuracy with precision 60% and recall 25%

Figure 8. Results.

VIII. OBSERVATIONS AND FUTURE WORK

A. Observations:

While applying the Manual Interpretation of being an ex-oplanet researcher, it is hard to search with small conveying signals, we observe that lots of exoplanet are undiscovered and visible through telescope so are just finding and clarifying an exoplanet with the help of Neural Network which is also call as Deep Learning, we will also be 1-D convolution for the betterment at at last ANN which is Artificial Neural Network for the better Accuracy, From this we will get to know the Recall and precision using the matrix and the various conditions True positive which means planet is there in real and we caught while using the Algorithm this are our Observations if not found it would be false negative. At last observing all this data we will get to know about undiscovered stars and Exoplanets.

B. Future Work:

We need to collect all required data from various possible sources to train the Artificial Intel-ligence. After gathering the data, we will start the implementation of our project by using the required algorithms and the data with reference to the research papers referred for this topic.

IX. PROJECT TIMELINE
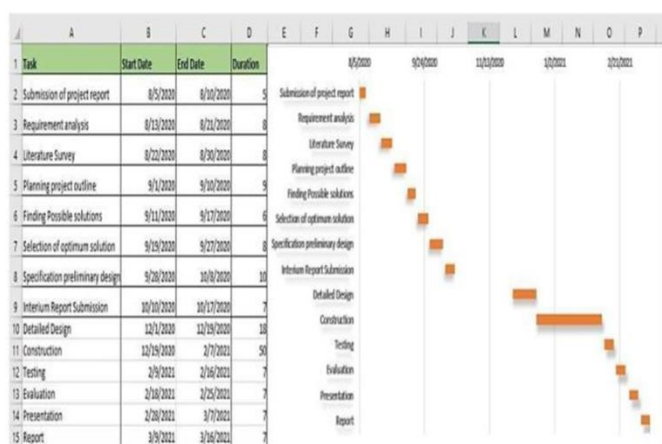
Figure 9. Project Timeline.

X. ACKNOWLEDGEMENT

XI. REFERENCES

[1]. Christopher J. Shallue and Andrew Vanderburg. "IDENTIFYING EXOPLANETS WITH DEEP LEARNING: A FIVE PLANET RESONANT CHAIN AROUND KEPLER-80 AND AN EIGHTH PLANET AROUND KEPLER-90" December 2014

[2]. Anne Dattilo, Andrew Vanderburg. "IDENTIFYING EXOPLANETS WITH DEEP LEARNING II: TWO NEW SUPER-EARTHS UNCOVERED BY A NEURAL NETWORK IN K2 DATA" June 2017

[3]. Kyle A. Pearson, Caitlin A. Griffith. "IDENTIFYING THE EXISTENCE OF EXO PLAN-ETS USING MACHINE LEARNING" May 2017

[4]. Nagesh Singh Chauhan. ". "SEARCHING FOR EX-OPLANETS— those planets beyond our own solar system - USING MACHINE LEARNING and implement these searches in Python" May 2019

Air Pollution Handling Using Machine Learning

Madan A[1], Banuprathap Reddy P[1], Dhanush G P[1]

[1]Department of Computer Science and Engineering, New Horizon College of Enginering, Bangalore, Karnataka,

India

ABSTRACT

The interpolation, prediction, and have analysis of fine-gained air quality area unit 3 vital topics within the space of urban air computing. The solutions to those topics will offer very helpful info to support pollution management, and consequently generate nice social and technical impacts. Most of the prevailing work solves the 3 issues one by one by completely different models. during this paper, we tend to propose a general and effective approach to resolve the 3 issues in one model referred to as the Deep Air Learning (DAL). the most plan of dekaliter lies in embedding feature choice and semi-supervised learning in numerous layers of the deep learning network. The planned approach utilizes knowledge} relating the untagged spatiotemporal data to enhance the performance of the interpolation and also the prediction, and performs feature choice and association analysis to reveal the most relevant options to the variation of the air quality.

Keywords— Deep Air Learning, Interpolation, Deep Learning, Machine Learning, Semi-Supervised Learning.

I. INTRODUCTION

This project is motivated to handle all challenges by pollution info contained within the untagged knowledge and performing arts feature choice and association analysis for the air connected knowledge. The interpolation, prediction, and have analysis of fine-gained air quality area unit 3 vital topics within the space of urban air computing. A good interpolation solves the matter that there area unit restricted air-quality-monitor-stations whose distribution is uneven during a city; an exact prediction provides valuable info to shield humans from being broken by air pollution; an inexpensive feature analysis reveals the most relevant factors to the variation of air quality. In general, the solutions to those topics will extract very helpful info to support

pollution management, and consequently generate nice social and technical impacts.

However, there exist many challenges for urban air computing because the connected knowledge have some special characteristics. First, since there area unit poor air-quality-monitor stations during a town because of the high value of building and maintaining such a station, it's costly to get tagged coaching samples once managing fine-gained air quality. Second, the tagged knowledge of the air-quality-monitor-stations area unit incomplete, and there exist innumerable missing labels of the historical knowledge in your time periods for a few stations. the rationale for the unfinished labels is said to the air quality monitor devices. In general, every station solely has one monitor device that must be maintained at intervals, so there'll be no outputs for the station once the device is being maintained,

recalibrated, or has alternative issues. Third, the forms of urban air connected knowledge area unit varied for the event of knowledge acquisition technologies. However, there's not a universally accepted judgment to reveal the most causes of the prevalence and dissipation of pollution, particularly the pollution of PM2.5. Hence, it's exhausting to understand that what forms of knowledge area unit the most relevant options for interpolation and prediction and also the key factors for atmosphere departments to forestall and management pollution.

This project is motivated to handle of these challenges by utilizing knowledge} contained within the untagged knowledge and also the spatio-temporal knowledge and performing arts feature choice and association analysis for the urban air connected data. tho' tagged knowledge area unit tough or costly to get, massive amounts of untagged examples will typically be gathered cheaply. In general, untagged knowledge will facilitate in providing info to raised exploit the geometric structure of the information. Moreover, most of the urban air connected knowledge contain each house and time info.

Fig. Health Effects Of Air Pollution

II. RELATED WORK

"Long-term inhalable particles and different air pollutants associated with mortality in non-smokers". long close concentrations of inhalable particles but ten microm in diameter (PM10) (1973- 1992) and different air pollutants-total suspended sulphates, dioxide, ozone (O3), and N dioxide-were associated with 1977-1992 mortality during a cohort of half-dozen,338 non-smoking California Seventh-day Adventists. In each sexes, PM10 showed a robust association with mortality for any mention of non-malignant disease on the death certificate, adjusting for a good vary of probably contradictory factors, together with activity and indoor sources of air pollutants. The adjusted relative risk (RR) for this explanation for death as related to AN interquartile vary (IQR) distinction of forty three d/yr once PM10 exceeded a hundred microg/m3 was one.18 (95% confidence interval [CI]: one.02, 1.36). In males, PM10 showed a robust association with carcinoma deaths-RR for AN IQR was two.38 (95% CI: one.42, 3.97). gas showed an excellent stronger association with carcinoma mortality for males with AN RR of four.19 (95% CI: one.81, 9.69) for the IQR distinction of 551 h/yr once O3 exceeded a hundred components per billion. dioxide showed sturdy associations with carcinoma mortality for each sexes. different pollutants showed weak or no association with mortality.

"Health risks of gas from long-range transboundary air pollution". gas could be a extremely aerophilous compound fashioned within the lower atmosphere from gases (originating to an outsized extent from phylogeny sources) by chemical science driven by radiation. because of its extremely reactive chemical properties, gas is harmful to vegetation, materials and human health. within the layer, gas is additionally AN economical greenhouse emission. This report summarizes the results of a multidisciplinary analysis progressing to assess the results of gas on health. The analysis indicates that gas pollution affects the health of most of the populations of Europe, resulting in a

good vary of health issues. the results embrace some twenty one 000 premature deaths annually in twenty five global organization countries on and when days with high gas levels. Current policies ar depleted to considerably cut back gas levels in Europe and their impact within the next decade.

"Long-term exposure to pollution and respiratory disease hospitalisations in older adults: a cohort study". Gliomas belong to a gaggle of central systema nervosum tumours and accommodates numerous sub-regions. Gold commonplace labelling of those sub-regions in picture taking imaging is important for each clinical and procedure studies, together with radiomic and radio genomic analyses. Towards this finish, we tend to unleash segmentation labels and radiomic options for all pre-operative multimodal resonance imaging (MRI) (n=243) of the multi-institutional brain tumor collections of The Cancer order Atlas (TCGA), in public on the market within the Cancer Imaging Archive (TCIA). Pre-operative scans were known in each brain tumor (TCGA-GBM, n=135) and low-grade-glioma (TCGA-LGG, n=108) collections via tomography assessment. The brain tumor sub-region labels were made by an automatic progressive methodology ANd manually revised by an skilled board-certified neuroradiologist. an in depth panel of radiomic options was extracted supported the manually-revised labels. This set of labels and options ought to modify i) direct utilization of the TCGA/TCIA brain tumor collections towards repeatable, duplicatable and comparative quantitative studies resulting in new prognostic, prognostic, and diagnostic assessments, still as ii) performance analysis of computer- assisted segmentation strategies, and comparison to our progressive methodology.

"Air pollution, pollens, and daily admissions for asthma". A study was undertaken to research the link between daily hospital admissions for respiratory disease and pollution in London in 1987-92 and therefore the doable contradictory and modifying effects of mobile spore. For all ages along and therefore the age teams 0-14, 15-64 and 65+ years,

Poisson regression was accustomed estimate the relative risk of daily respiratory disease admissions related to changes in gas, dioxide, dioxide and particles (black smoke), dominant for time trends, seasonal factors, calendar effects, contagion epidemics, temperature, humidity, and autocorrelation. freelance effects of individual waste products and interactions with aeroallergens were explored victimisation 2 pollutant models and models together with spore counts (grass, oak and birch). In all-year analyses gas was considerably related to admissions within the 15-64 people (10 ppb eight hour gas, 3.93% increase), dioxide within the 0-14 and 65+ age teams (10 ppb twenty four hour dioxide, 1.25% and 2.96%, respectively), dioxide within the 0-14 people (10 micrograms/m3 twenty four hour dioxide, 1.64%), and black smoke within the sixty fifth people (10 micrograms/m3 black smoke, 5.60%). vital seasonal variations were discovered for gas within the 0-14 and 15-64 age teams, and within the 0-14 people there have been negative associations with gas within the cool season. In general, accumulative lags of up to 3 days attended show stronger and a lot of vital effects than single day lags. In two- waste product models these associations were most sturdy for gas and least for dioxide. There was no proof that the associations with air pollutants were because of contradictory by any of the pollens, and tiny proof of AN interaction between spores and pollution apart from synergism of dioxide and grass pollen in youngsters ($p < 0.01$).

III. METHODOLOGY

This project is driven to handle of these challenges by utilizing the knowledge contained within the unlabeled information and also the information and playing feature choice for the urban air connected data. the tagged information square measure tough or overpriced to get, massive amounts of unlabeled examples will usually be gathered cheaply.Here, our system solves all existing drawback exploitation single model. With higher rate of accuracy within the

prediction of air quality. The planned feature choice and analysis technique reveals the importance of various input options to the predictions of the neural networks, so has the flexibility to reveal some inner mechanism of the black-box deep models, that doesn't limit to pollution hindrance and management, however may be applied to several different applications, like diagnosis and coercion detection. for instance, one cannot take actions supported a model's predictions while not knowing whether or not the black-box model is trustworthy or not, otherwise the implications is also harmful.
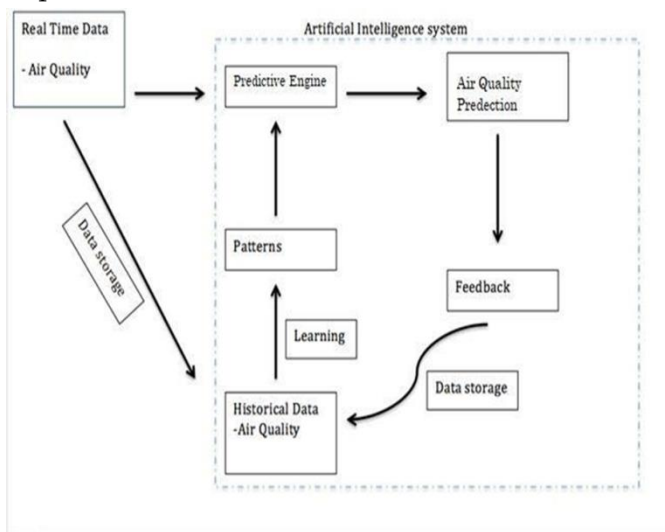
Fig. System Aarchitecture

A data flow diagram (DFD) is graphic representation of the "flow" of data through an information system. A data flow diagram can also be used for the visualization of data processing (structured design). It is a common practice for designers to draw a context level DFD first which shows the interaction between the system and outside entities. DFD's show the flow of data from external entities into the system, how the data moves from one process to another, as well as its logical storage.
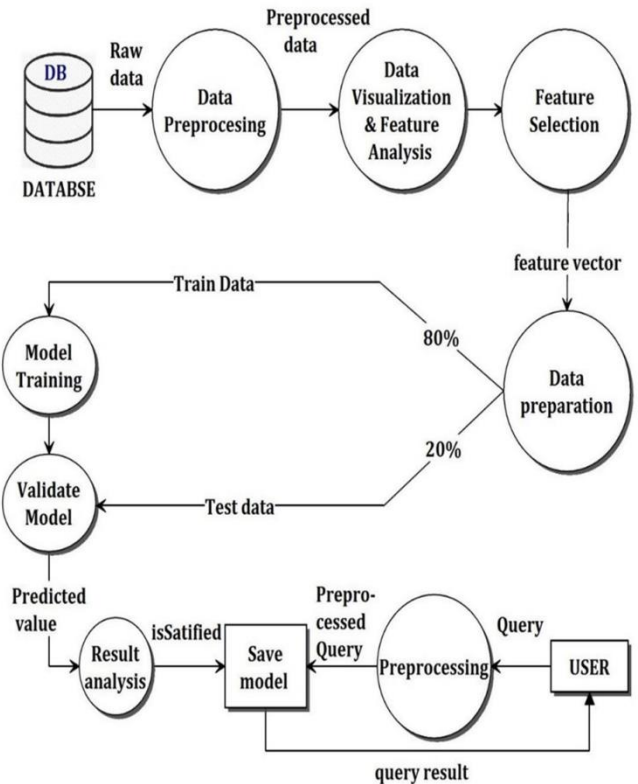
Fig. Data flow diagram.

IV. CONCLUSION AND FUTURE WORK

The project work has been disbursed to reinforce the accuracy of prediction by victimization Deep air learning techniques. During this approach, information square measure pre-processed before any mining techniques square measure applied within the discovery of information. The whole method would be laid low with the presence of orthogonal, redundant and insignificant options within the high dimensional information. These options have an effect on the accuracy of prediction, clustering, classification, regression. Therefore, there arises a requirement to reinforce accuracy of the prediction and different learning approaches.

The field of Deep air learning Analysis is an exciting new analysis direction because of sizable amount of real-world applications wherever discovering people's opinion is very important in higher decision-making.

The event of techniques for the document-level sentiment analysis is one in every of the many parts of this space. Recently, folks have started expressing their opinions on the {online|the net} that accrued the requirement of analysing the narrow online content for numerous real-world applications. Heaps of analysis is gift in literature for detective work sentiment from the text. Still, there's an enormous scope of improvement of those existing models. Existing sentiment analysis models is improved more and common- sense data.

Further upgrades to our work which could be performed are:

1]. In providing crucial information to support Air Pollution Control.

2]. In generation of great societal aware and technical impact due to air pollution.

3]. Due to time constraint, the project has been restricted to a sample, but in future, People can build the big model of it, which will also improve the accuracy of prediction.

4]. The accuracy we could achieve was around 96%. In future, this accuracy can further be increased by performing advanced operations.

5]. Database can be implemented.

6]. Extra features can be added where a user can upload the data set directly and the accuracy would be predicted dynamically.

V. REFERENCES

[1]. R. M. GRAY, "VECTOR QUANTIZATION," IEEE ASSPMAGAZINE, PP. 4–29, APRIL 1984.

[2]. C.-C. CHANG, T. S. NGUYEN, AND C.-C. LIN, "A REVERSIBLE COMPRESSION CODE HIDING USING SOC AND SMVQ INDICES," INFORMATION SCIENCES, VOL. 300, PP. 85–99, 2015.

[3]. H.-T. HU, L.-Y. HSU, AND H.-H. CHOU, "VARIABLE-DIMENSIONAL VECTOR MODULATION FOR PERCEPTUAL-BASED DWT BLIND AUDIO WATERMARKING WITH ADJUSTABLE PAYLOAD CAPACITY," DIGITAL SIGNAL PROCESSING, VOL. 31, PP. 115–123, 2014.

[4]. N. FARVARDIN, "A STUDY OF VECTOR QUANTIZATION FOR NOISY CHANNELS," IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 36, NO. 4, PP. 799–809, 1990.

[5]. N. FARVARDIN AND V. VAISHAMPAYAN, "ON THE PERFORMANCE AND COMPLEXITY OF CHANNEL-OPTIMIZED VECTOR QUANTIZERS," IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 37, NO. 1, PP. 155–160, 1991.

[6]. Y. LINDE, A. BUZO, AND R. M. GRAY, "AN ALGORITHM FOR VECTOR QUANTIZER DESIGN," IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 28, NO. 1, PP. 84–95, 1980.

[7]. R. EBERHART AND J. KENNEDY, "A NEW OPTIMIZER USING PARTICLE SWARM THEORY," MHS'95. PROCEEDINGS OF THE SIXTH INTERNATIONAL SYMPOSIUM ON MICRO MACHINE AND HUMAN SCIENCE, PP. 39–43, 1995.

[8]. "PARTICLE SWARM OPTIMIZATION," IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORK, VOL. 4, PP. 1942–1948, 1995.

Fingerprint Recognition using Deep Learning

Yogitha[1], Kushala P[2], A Rishitha Reddy[2], Bhoomika G[2]

[1]Assistant Professor, Department of Computer Science & Engineering, New Horizon college of Engineering, Bangalore, Karnataka, India

[2]Department of Computer Science & Engineering, New Horizon college of Engineering, Bangalore, Karnataka, India

ABSTRACT

Validity and consistency of fingerprint recognition has proven to be one of the most reliable methods for human identification. The fingerprint matching issue is conceived as an arrangement in which a model is created to learn to distinguish between a true and impostor pair of fingerprints. Previously, they used to exercise feature extraction prior to comparing a pair of fingerprints. Also, recently CNN has presented marvelous success for many images processing task. However, there are only a couple of attempts to develop a complete CNN method to influence challenges in the fingerprint recognition problem. We attempted to build a CNN-based fingerprint matching system in this research. The ability to learn fingerprint patterns directly from raw pixels in photos is a significant contribution of the technology. Incomplete and partial pairs of fingerprints were considered for feature extraction in order to achieve resilience and characterize commonalities broadly.

Keywords — MobileNet v1, ReLu Activation, Convolutional Neural Network, Deep Learning

I. INTRODUCTION

Fingerprint is the impression the finger leaves on any surface. It is created due to the friction between the surface and the papillary edges of the finger. Fingerprints are highly unique and have distinguishing features such as grooves and edges that are differently arranged in each individual. This consistency and uniqueness served as an inspiration to use these small but efficient features for human recognition, verification and security purposes. Fingerprint recognition is an automated way of identifying an individual by processing the fingerprint and comparing the extracted features with

another fingerprint. This procedure has become popular due to its reliable uniqueness and ease of acquisition.

With development in technology, Biometrics was introduced as an automated way of establishing the identity of a person. Fingerprint recognition provided the basis for this technique.

Fingerprint matching takes in only the required patterns of a fingerprint. Fingerprinting serves as a base for the first forensic professional organization formed, the International Association for Identification (IAI), in 1915. The distinctiveness of the fingerprints made it the most widely used biometric identifier in the 2000s. For authentication

reasons, recognition algorithms compare previously saved fingerprint templates to an individual's fingerprints. To do so, either the original image and the individual's image must be directly compared, or certain attributes must be compared. The uniqueness and difficulty of altering the fingerprints of a human and the durability over the life makes them suitable as long-term markers of human identity.

II. RELATED WORK

Biometric Recognition Using Deep Learning: A Survey

Deep learning-based models have been masterful in accomplishing results in computer vision, speech recognition and natural language processing tasks in the last few years. These models have been an essential for steering the ever-increasing scale of biometric recognition problems, from cellphone authentication to airport security systems. Deep learning-based models have progressively been leveraged to enhance the accuracy of different biometric recognition systems over the last few years. For each biometric system, we first introduce the available datasets that are widely used in the literature and their characteristics. Biometric features hold a unique place when it comes to recognition, authentication, and security applications. They won't be able to get lost, unlike token-based features such as keys and ID cards and they cannot be forgotten. Furthermore, it is almost impossible to perfectly imitate or duplicate a fingerprint. Even though recently there have been attempts to generate and forge various biometric features, there have also been methods proposed to distinguish fake biometric features from authentic ones. Fingerprint constitutes ridges, grooves and valleys, forming unique shapes and patterns. Fingerprints have major local portions called Minutiae which can be used to determine the uniqueness of the fingerprint, it has two most important ones that are: ridge endings and ridge

bifurcations. The summary of the recent deep learning-based models (till 2019) for biometric recognition. As opposed to the other surveys, it provides an overview of most used biometrics. Deep neural models have demonstrated better and efficient working over older models for various biometrics.

Comparison of Deep Learning Model for Biometrics based Mobile User Authentications

In this paper, Narsi Reddy, Ajita Rattani and Reza Derakhshani describe how deep learning technique provides developments in many applications like image identification, segmentation and other detections. They have elaborated how CNN is a class under artificial neural networks generally used in analyzing visual images, like the ones in camera-based mobile biometrics. We learned how the CNN method employs an input and output layer, as well as numerous hidden layers that include convolutional layers, pooling layers, fully connected layers, and normalizing layers. They compared several CNN architectures like VCG, ResNet, MobileNet V1 and V2, DenseNet and then proposed a model inspired by Mobile Net-v1. However, Mobile Net-v1 intrigued us for our study. They have briefly explained the popularity of Mobile Net-v1 deep learning architecture among mobile-centric They explained that the basic idea behind MobileNet is that instead of utilising standard convolution filters, they use convolution filters of size 3x3, the operation uses a 3x3 convolution filters split into depth-wise separable, followed by convolutions layer having resolution of 1x1. As a regular convolution, this provides the same type of filtering and combination technique. The new architecture, on the other hand, employs fewer stages and parameters.

Fingerprint Recognition Algorithm by Farah Dhib Tatar

In this paper, Farah Dhib Tatar mentions that the advantage of biometric identification is that each individual has their own physical characteristics that distinguishes them from others. This is what makes this type of recognition more reliable. She says that the techniques of fingerprint recognition are enormous and diversified, and are generally based on generic algorithms and tools. The article proposes a fingerprint recognition chain based of Filtering algorithms. The outputs for these algorithms are retrieved and validated using MATLAB. The results acquired are directly connected to two main criteria: the captured image quality and the processor used to process the images. There are many types of sensors used for image acquisition the widely used sensors in the market are the CMOS sensors as they allow to decrease the overall price of cameras, since they possess all the elements needed for the composition of cameras. With regard to the implementation of code, there are various categories of processors that can be used ranging from those provided by companies specialized in embedded manufacturing such as Altera, Xilinx, Texas Instrument etc. Or "free" processors such as Raspberry Pi, Beaglebone, Arduino, etcetera. The software performance and the code remain strongly dependent on these two steps and changes mainly in accordance to the types of processors used for the processing of the image.

III. ALGORITHM

- Convolutional Neural Networks (CNN) has proven to be effective at image classification techniques, that includes classic problems like handwritten digit recognition. The CIFAR-10 and CIFAR-100 algorithms provide state-of-the-art solutions to real-world problems like facial recognition, pose estimation, gray-scale image colorization, and many others.

- Each convolutional layer is made up of a bank of filters (also known as weights) that are convolved with the preceding layer's output or the input image if it is the first layer to produce a reaction.

- The max-pooling layer selects the maximum value in a MxM window to conduct sub-sampling on the outputs provided by the previous convolutional layer or layers. All max-pooling layers in this work are 2x2 with a stride of 2, lowering the output size by 2.

- Fully Connected Layer: This layer connects a group of neurons to each of the previous layer's neurons.
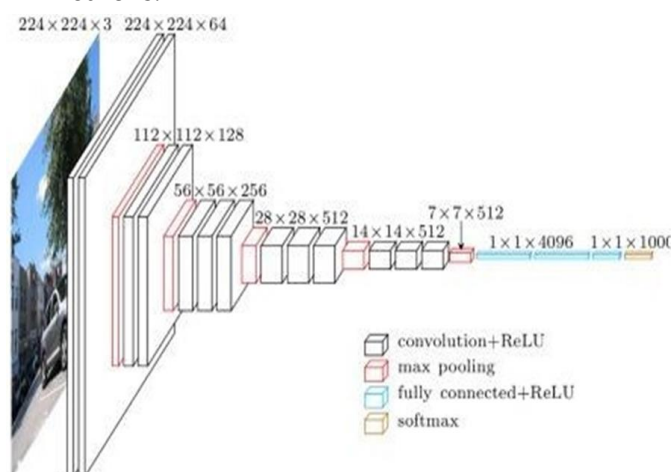
FIG 2.1 CNN ARCHITECTURE

IV. METHODOLOGY

This Project, a setup of Keras, which is an open source Deep Neural Network system with the TensorFlow backend which is utilized to assemble and prepare the Convolutional Neural Network. In the frontend development of this study, for a framework that employs the Python programming language, we created an MVC. This system includes a generator that creates MVC folder structures quickly and easily, as well as the Bootstrap framework and the fact that it is open source. At backend, we have used google collaborator for running the python code, as it gives quicker TPU and RAM size of 12gb for the quicker preparing measure.

In our study of fingerprint recognition, we have collected a dataset containing 600 unique fingerprints that belongs to 60 different people i.e., 10 fingers

from each individual along with their personal details like name, age, address, etc. We have taken this dataset from Kaggle. On implementing the Mobile net-v1 architecture of CNN algorithm in our study, we split the data into train set (having 80% of data) and test set (having 20% of data). In the CNN model algorithm generation, firstly we have performed data cleaning, then implemented the Convolution Layer, where each layer has a set of filters that are convolved with the previous layer's output or the input image if it's the initial layer to produce a response. For the first surface, we have set filter to 32 and eventually increased it up to 128 at third layer in the Conv2D() function. We have used layer activation as a parameter of Conv2D function, the activation used is ReLu which applies the rectified linear unit activation function. Then we have implemented Batch Normalization to stabilize the layer. Then we used MaxPooling2D(), which reduced the output size by 2 by making all max-pooling layers 2X2 with a stride of 2.

FIG 3.1 MODEL SUMMARY

In the training model, we have set the epoch value to 5 and noticed an increasing accuracy as expected.

FIG 3.2 TRAINING RESULT

On subjecting all the above implementations, a graph is plotted depicting the training and validation accuracy of the subject's fingerprint ID recognized successfully and the loss incurred.
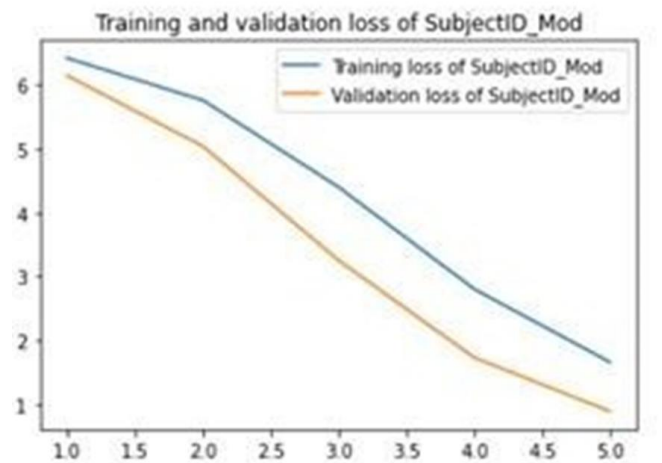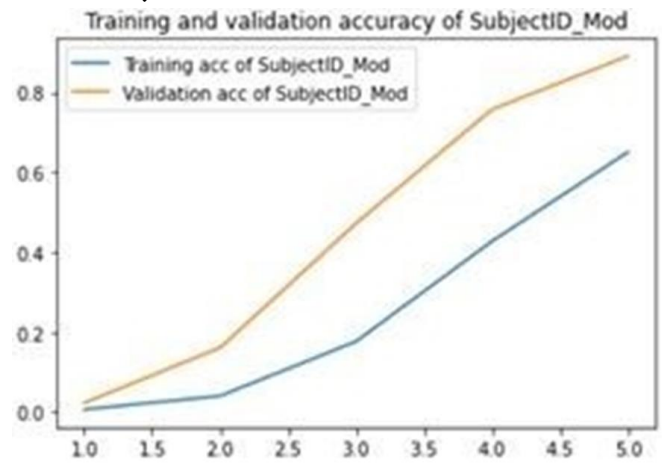
FIG 3.3 GRAPH OF ACCURACY DEVELOPED AFTER TRAINING

V. CONCLUSION

Many features in a human body can be considered as a unique identifier in bio metrics. We have proposed a unique mark acknowledgement in which we venture for separating fingerprint features before implementing CNN classification. The model is supposed to identify a fingerprint image, recognize the print, clean the pattern and differentiate the grooves and edges in the fingerprint. Then, load each fingerprint which is unique to the system along with the details as to whom it belongs to so that it can be later searched and retrieved from the system when an image is uploaded to the system.

The model is supposed to identify a fingerprint image, recognize the print, clean the pattern and differentiate the grooves and edges in the fingerprint. Then, load each fingerprint which is unique to the system along with the details as to whom it belongs to so that it can be later searched and retrieved from the system when an image is uploaded to the system.

VI. REFERENCES

[1]. Biometric Recognition Using Deep Learning: A Survey by Hang Su and David Zhang.

[2]. Comparison of Deep Learning Model for Biometrics based Mobile User Authentications by Narssi Reddy, Ajitha Ratani and Reza Deraakhshani

[3]. Fingerprint Recognition Algorithm by Farah Dhib Tatar.

Signature Originality Check using Variation Extraction

Samruddhi Chavan[1], Hitesh Choudhari[1], Vishwajit Jadhav[1], Rakesh Suryawanshi[1]

[1]Department of Computer Engineering, University of Mumbai, A. C. Patil College of Engineering, Navi Mumbai, India

ABSTRACT

In today's world, handwritten signature plays vital role in authentication. Institutions like banks, government offices are using the handwritten signatures for authentication. As signature is considered as an important parameter in authentication, need to verify them is proportionally increasing. Hence, we always required an intelligent signature verification system. There are several intra-personnel variations for the same person's signatures that can be analysed in order to validate the signature. As a result, we propose the scale and rotation unvarying technique, which entails extracting a rotation unvarying sub uniform as a feature vector from an image and using it for verification. The Local Binary Pattern (LBP)in signature verification has used to obtain the dissimilarity patterns. Chi square distance is used for calculating the variation factor. Also, Center of Mass (COM) and Center of Gravity (COG) of a signature is calculated and used for verification of signature.

Keywords — Intra-personal variation, Local Binary Pattern, KNN, KSVM, Variation Extraction, Off-line signature verification.

I. INTRODUCTION

Handwritten Signature Recognition is an important behavioral biometric which is used for wide range of identification and authentication applications.

Our proposed system deals with variations present in signature for its originality check. Variations are the dissimilarities in the signatures of a particular author. Local binary pattern is used to extract the feature vector of the signature. This feature vector is in the form of histogram. Then, chi square distance method is used to calculate the dissimilarities between two histograms. COG and COM is also calculated of the signature. Finally, the chi square distance, COG and COM have been used to train the model and conclude whether the signature is forged or genuine.

II. LITERATURE SURVEY

Signature Verification systems have been implemented earlier using different technologies like Vector Machines, Hid- den Markov Models (HMM), Wavelet transforms and Fourier transforms.

There are some biometric verification systems which are using signature characteristics, discrete Fourier transform and Wavelet transform for the signature feature extraction. Such system extracts the feature not only from the entire signature but also from its individual parts. These extracted features along with its statistical characteristics like speed, magnitude and pressure are also considered classifying the signature as forged or genuine. [2]

In an Automatic Signature Verification, questioned signatures are usually compared with source

signatures. A writer independent system requires a set of reference signatures from various authors in order to develop the model of the system. This method addresses the problem of an automatic signature verification when no reference signatures are available. [3]

In this paper, researchers have contributed several techniques and ideas for signature verification. It was examined that the models were not efficient. Thus, a GPU based offline signature verification model had been proposed. Faster response can be achieved using GPU based model. The model is evaluated on the basis of parameters namely: FAR and FRR. [1]

III. PROPOSED SYSTEM

A. Architecture

The proposed system is divided into several modules which are supposed to work in accordance with one another. The architecture is shown in figure 1.

The modularity of the systems is represented by following modules:

- Pre-processing:

In pre-processing stage, the image will be converted to binary image and all images will be resized to same dimensions. Also, COG and COM will be calculated and stored in knowledge base. [7]

- Variation Extraction:

Here the dissimilarities will be extracted from the images. Local binary pattern (LBP) will convert image to histogram. Finally, chi square distance will be used to extract the dissimilarities from the histograms. [4], [6]
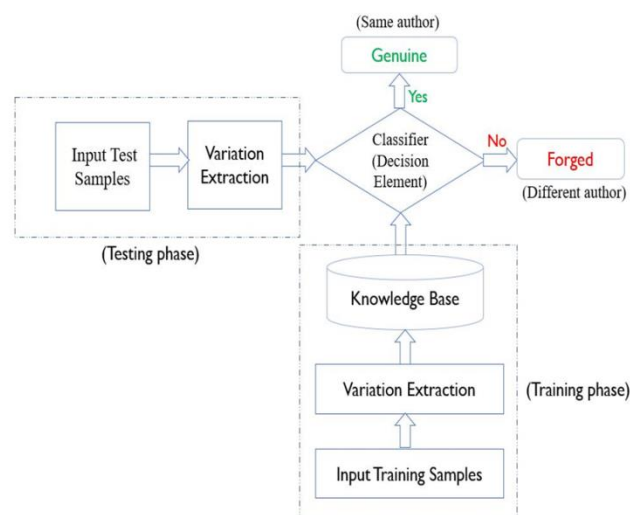
Fig. 1 Architecture

- Knowledge Base:

The histogram generated using LBP and the chi square distance between those histograms will be stored in knowledge base.

- Classifier:

This module will be used to study the dissimilarities obtained, and using those will be able to conclude whether the signature is forged or genuine.

B. Pre-processing

1) RGB to Gray Conversion: Signature image is captured either by using the digital camera or scanner. The captured signature is in RGB scale and is converted to gray-scale using the following equation:

GrayScale = $(0.2989 * R + 0.5870 * G + 0.1140 * B)$ (1)

where R,G and B are the RGB color components of the input image

2) Image Binarization: Image binarization is the process of converting a gray-scale image to black-and-white image. It metamorphoses all the 256 gray shades to just two shades i.e. black and white. This is also known as thresholding. For this we will use Otsu algorithm. [7]

Otsu Algorithm:

- Conversion of RGB image to gray-scale.

- Calculate image histogram.
- If the intensity of a region is high, then replace those pixels with white and replace with black in opposite case

The below figure 2 shows image binarization.
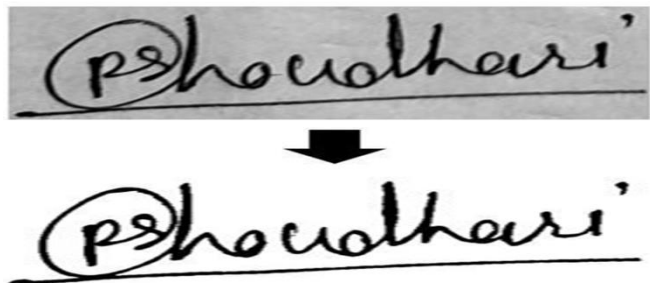
Fig. 2 Image Binarization

3) Center of Gravity: Center Of Gravity supplies the information about pixel's density. It divides the entire signature image at a specific point G(x,y) in a such way that image divided into vertical and horizontal regions which will have same number of pixels. [5]

4) Center of Mass: Center Of Mass is similar to centroid of images. Here, we have calculated COM by using regionprops library of python whose important property is COM. Centroid of each object present in an binary image is calculated where objects can be continuous and discontinuous and is displayed using regionprops. In our case, signature is a continuous object.

5) Resizing: The signature images are re-scaled to a (500 px,300 px) size using the image resize option. Image resizing is used when you need to increase or decrease number of pixels in an image.

C. Variation Extraction

This stage will be executed using Local binary pattern and chi square distance.

1) Local Binary Pattern: Local Binary Pattern (LBP) is efficient texture operator which replaces values of the pixels of an image by thresholding the neighbourhood of each pixel and considers the pixel value as a binary number. To perform LBP, first the image is converted to gray-scale format.

Then a center pixel is selected and a radius of value r is defined surrounding a center pixel. Then, LBP value is calculated for that center pixel by thresholding the neighbouring pixels. As shown in this image. The figure 3 below shows this procedure:
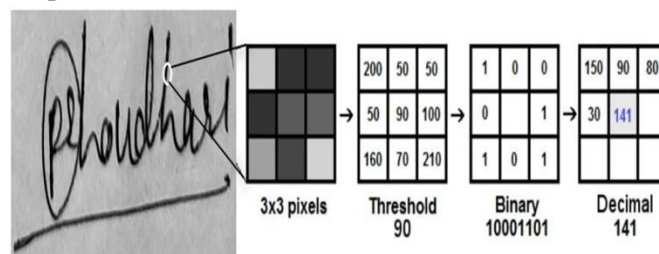
Fig. 3 LBP Working Steps

As compared to other texture-oriented images like facial or materials, signatures are more edge oriented. Thus, to extract detailed variations, we have converted gray-scale image to binary image and performed LBP on the same.

2) Histograms Generation: After LBP image is obtained as shown in above figure 3 in the last step, the entire image is divided in grids along X and Y axes, as can be seen in the following image:
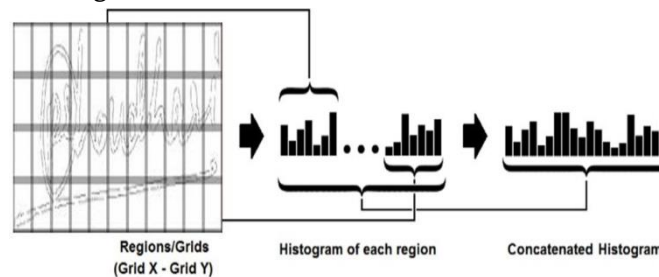
Fig. 3 Histogram Generation

Based on the figure 4 above, following procedure is carried out:

- Each histogram (from each grid) will be generated, rep- resenting the instances of each pixel intensity.
- Then, integrate each histogram to create one bigger histogram. The final histogram represents the characteristics of the original image.

We need to extract feature which are rotation invariant as many signatures have slope therefore, to do so we have used Sub uniform local binary pattern. To achieve rotation invariant the maximum value bin in each grouped bins is shifted to the farthest left and other bins are circularly shifted.

3) *Chi-Square Distance:* Chi-square histogram distance method can be used to find dissimilarity between two histograms. It has been used in various image processing techniques like detection, classification. While calculating the difference between two histograms, other methods gives less importance to variation between larger bins as compared to variation between smaller bins, that should be avoided. The chi-square distance method takes this into account. [6]

Chi Square Distance $= \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{(x_i + y_i)}$ (2)

D. Classification

In machine learning, a classifier is an algorithm that classifies data into one or more classes. In our system we have used KNN and KSVM classification models.

An output is always class dependent in case of KNN classification. An object assigned to a class depending on its neighbour's class. For eg. If an object is surrounded by k nearest neighbours (where k is an integer) belonging to class "Genuine", then it will be automatically assigned to "Genuine" class.

Kernel machine algorithms are extensively used in pattern scanning and one of the best kernel algorithms is SVM. SVM algorithm creates a fence between two classes and depending upon the number of objects of a class on the fence, this classifier provides the result. This algorithm is used in case of 2 dimensions, otherwise KSVM is used.

IV. IMPLEMENTATION

A. Training
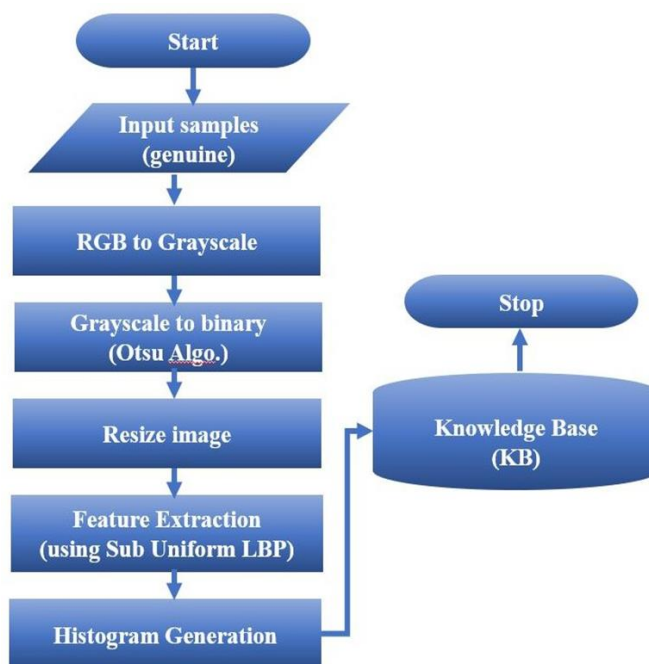
Fig. 4 Flowchart A

Flowchart A represents the procedure to be performed on reference (genuine) images. Total 20 images were used.

Step 1 : RGB image converted to Grayscale image.
Step 2 : Grayscale image converted to binary image.
Step 3 : Resizing the image.(size (500x300))
Step 4 : Feature extraction using LBP and Generation of Histogram.
Step 5 : Saving histograms in knowledge base.

Here, 20 reference histograms have been generated. Flowchart B represents the procedure to be performed on training (genuine and forged) images. Total 100 images (out of which 70 are forged and 30 are genuine) have been used.
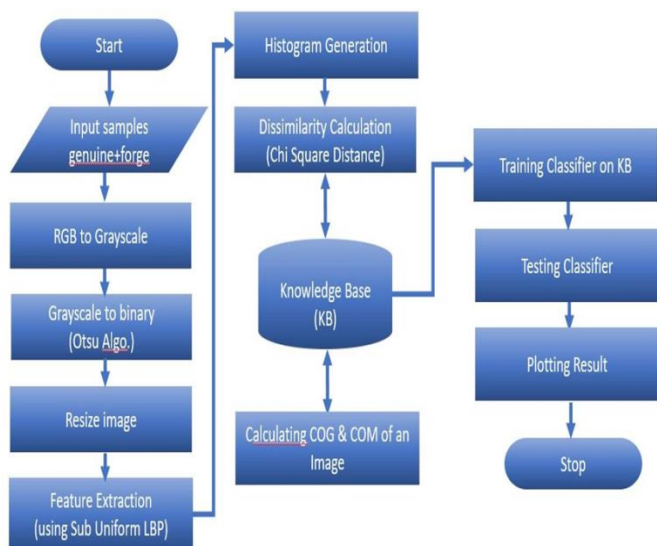
Fig. 5 Flowchart B

Step 1 - 5: Pre-processing steps (same as reference image).

Step 6 : Chi-square distance between each training image's histogram and all reference histograms is calculated i.e. for each training image, 20 chi-square distances have been stored in knowledge base. Thus, dataset of 2000 records have been generated.

Step 7 : COM and COG for each training image is calculated and stored in knowledge base, adjacent to its corresponding training image's Chi-square distance.

Step 8 : 80% of total 2000 entries are used for training the classifier and remaining 20% for testing.

Step 9 : Finally, the trained and tested classifier has been used for determining forged and genuine signature.

B. Testing:

The proposed model has been tested on a single signa- ture image at a time. All the pre-processing steps (RGB to grayscale conversion up to histogram generation) have been carried out. Then Chi-square distance between given image's histogram and all the 20 reference histograms have been calculated and stored in knowledge base. Also, COG and COM for the same has been calculated and stored accordingly

Further, the trained model has been used to predict whether the given signature was "Forged" or "Genuine".

V. RESULT AND ANALYSIS

The experiment was divided in two phases, in first phase we considered only single parameter i.e. Chi-square distance and the second where we added two extra parameters namely COM and COG. After adding these two parameters we observed following changes in accuracy as shown in the Table 1 below.

Dataset Size	Algorithm	Accuracy(%)
Before Adding Extra Parameters		
2000	KNN	85
	KSVM	76.7
After Adding Extra Parameters		
2000	KNN	99.5
	KSVM	96.2

Table I Addition of Parameters

During the analysis of results, it has been observed that all the three parameters have equal contributed. Out off total 100 training images 70 were forged. Those 70 images were forged by multiple authors in order to gather variety of signatures to extract all possible variation patterns. As our model is writer dependent, the above table represents data of a single author. Thus, to check the consistency, proposed model has been tested on four more authors and the results obtained of all the 5 authors are shown in Table 2 below:

Author	Algorithm	FAR	FRR	Accuracy (%)
A1	KNN	0.003	0.008	99.50
	KSVM	0.03	0.05	96.25
A2	KNN	0	0	100
	KSVM	0.003	0.007	99.5
A3	KNN	0.01	0	98.75
	KSVM	0	0	100
A4	KNN	0.02	0.03	97.75
	KSVM	0.04	0.05	95.50
A5	KNN	0	0	100
	KSVM	0	0	100

Table II Results of Multiple Authors

Following are the equations for calculating FAR and FRR,

$$\text{FAR (False Acceptance Rate)} = \frac{FP}{FP+TP} \quad (3)$$

$$\text{FRR (False Rejection Rate)} = \frac{FN}{FP+TP} \quad (4)$$

where,

FP: False positive, FN: False Negative, TN: True Negative, TP: True Positive

VI. CONCLUSION

The study proves that the most promising methods for the variation extraction are the Local Binary Pattern and Chi-square distance and the use of parameters like COG and COM, because the accuracy on the method of Kernel Support Vector Machine and k-Nearest neighbors maximum in relation to others.

This model further can be developed as a writer independent model i.e. one model can be used to determine the authenticity of a signature for multiple authors.

VII.   ACKNOWLEDGMENT

The authors are grateful to A. C. Patil College of Engineering for its generosity and extending its utmost support and cooperation in providing all the provision of the work. We would like to express our heartfelt thanks towards all the team members for their sustained help.

VIII. REFERENCES

[1]. Saroj Kumar Chandra Amit Kumar Kar and Manish Kumar Bajpai. Parallel GPU based Offline Signature Verification Model. 2019.

[2]. Anna Epishkina Anastasia Beresneva and Darina Shingalova. Handwritten Signature Attributes

[3]. Moises Diaz. Investigating the Common Authorship of Signatures by Off-line Automatic Signature Verification without the Use of Reference Signatures. 2019.

[4]. Vivek H. Mahale. Image Inconsistency Detection Using Local Binary Pattern (LBP). 2017

[5]. Piotr Porwik and Tomasz Para. Some Handwritten Signature Parameters in Biometric Recognition Process. 2007.

[6]. Xiaopan Chen Fengbin Zheng 1 Wei Yang, Luhui Xu and Yang Liu. Chi-Squared Distance Metric Learning for Histogram Data. 2015.

[7]. Jamileh Yousef. Image Binarization using Otsu Thresholding Algorithm Algorithm. 2011

Medical Chatbot

Naymathula Khan[1], Syed Basith Hussain[1], Srikanth K S[1], Dr. S Sridevi[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Road, Panattur post, Kadubeesanahalli, Bengaluru, Karnataka, India

ABSTRACT

Artificial intelligence is now-a-days giving a new way for the medicine field. Therefore, only used by physicians for the use of prediction of diagnosis. However, this all systems are widely used, example, in the predicting of diseases and cancer. In the current studies, we have created which is capable of interacting with the patients (visual practitioner) through a speech recognition and word processing process and thus be able to interact independently with the patient, most importantly, for example, rural areas, where access to basic medical care is limited. In addition, the system will always be giving easy measurements of the T2DM potential interpretation of a given patient. In addition to the development of AI, we have continued to analyze the adoption of new AI in health care to estimate the effect of this program in the upcoming future.

Keywords : chatbot, forecasting, doctor, nlp, machine learning, AI

I. INTRODUCTION

Every year many peoples are treated in the but many patients are of non-emergency cases.

Compelling many hospitals to allow medical staff where ever they are not necessarily needed, creation of high staffing and emergency medical care for patients. In addition, it results in shortage of doctors in village areas, which can then lead to poorly taking cared for the patients, and the growing number of old peoples and population change, one of the best solutions to reduce this problem in future is to use AI in Medicine field. In addition, voice and speech identifying systems are used to support physicians in their everyday routine in hospitals, example., providing computer assisted documents, that is, they use speech and voice recognition to translate voice to text and prescribe patient results to complement health records. To date, however, speech and voice recognition has not found way Amazon into medical decision support systems. specialize systems used for the prediction of treatment options, This desired transformation will also come about in an effort to divide a sector that involve the distribution of many common presentations and a building of bigdata acquisitions and managements, development and the evaluation methods. AI can also improve the treatment or diagnosis, e.g., HIV drug resistance (Riemenschneider et al., 2016), breast cancer prediction (Montazeri et al., 2016), or type 2 disease mellitus (T2DM) (Talaei-Khoei and Wilson, 2018), however, these types cannot be used automatically and unattended. For example, T2DM predictive models studied more widely e.g., Perveena et al evaluated the performance of a T2DM predictor based upon affective students.

The authors demonstrate of an ability to predict the logistic regression models based upon Hba1c and

225

adiponectin or Hba1c and BMI. Chen & Pan (2018) used promotional algorithms for the prediction of T2DM. In contrast, Lall et al (2017) analyzed the genetic associations within genome-- wide relationship studies to determine the risk of disease. In the other recent studies, the data from the 14,000 smartwatch users have been analyzed and it is shown that infections can be predicted using smartwatch capabilities to perform the cardiogram. These observations rely heavily on the heart research by Tison et al. However, much of an AI system are used to analyze patient's data and make predictions about the potential solutions of treatment or to provide additional diagnosis or predictive data.

The authors demonstrate of an ability to predict the logistic regression models based upon Hba1c and adiponectin or Hba1c and BMI. Chen & Pan (2018) used promotional algorithms for the prediction of T2DM. In contrast, Lall et al (2017) analyzed the genetic associations within genome-- wide relationship studies to determine the risk of disease

In the other recent studies, the data from the 14,000 smartwatch users have been analyzed and it is shown that infections can be predicted using smartwatch capabilities to perform the cardiogram. These observations rely heavily on the heart research by Tison et al. However, much of an AI system are used to analyze patient's data and make predictions about the potential solutions of treatment or to provide additional diagnosis or predictive data. These types are usually based on mathematical and mechanical learning methods; however, they are not associated with speech and voice recognition or association for engaging with the patient. The objectives of the current study were the development of an AI communication with the patient real doctor, and demonstrating the usefulness in the automatic predictions of a model disease called T2DM, in a big patient group, and the design and the analysis of questionnaires at admission of AI in the medicine for young adults. The development of an AI is also dealing with the problems of mysterious district

(Mori et al., 2012), i.e., AI that mimics human behavior evokes the fears and frustrations of some viewers. However, there are already some of the researches suggesting that these problems can be genetic, as very young people who are heavily uses for an AI may have little chance of exposure (Hanson et al., 2005). Therefore, we all focuses on young adults in the current study to analyze the acceptance and the purpose of using an AI in medicine.

MOTIVATION BEHIND THE RESEARCH

Predicting disease using a data mine is one of the most exciting and challenging tasks. The shortage of specialists and high incorrectly obtained cases has necessitated the development of a quick and effective detection system. The main purpose of this work is to identify important patterns or features from medical data using the isolation. model and NLP. Qualities that are closely related to disease diagnosis can be identified. This will help doctors to understand the causes of disease more deeply.

II.   LITERATURE SURVEY

Literature survey is the most important step in the software development process (4). Machine learning technology adapts and responds to data, learning over time to better answer search queries. This helps provide better accuracy in search results at a speed beyond human capabilities

Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19

Tele medicine used by medical personnel to communicate with the patients during the recent Coronavirus outbreak, while trying to reduce the transmission of COVID-19 between patients and clinicians. In the midst of this epidemic, Tele medicine have the potential to help people by allowing patients to receive all the supportive care necessary without visiting hospital physically through their intellectual-based consultation program.

Therefore, telehealth will quickly and completely transform personal care into remote patient consultation. As a result, it developed the Multilanguage Conversational Bot based on AI using Natural Language Processing (NLP) to provide health care education for free, information and counseling to patients. The studies introduce an automatic novel system that acts as a visually impaired physician and is extensively trained for the purpose to communicate with human like patients. This application is purely based on uninterrupted design and integrates medical services by providing necessary preventive measures, home remedies, counselling sessions, health care tips, and symptoms which cover the most common diseases in villages of India. This paper proposes a discussion bot "Aapka Chikitsak" in the Google Cloud Platform (GCP) to bring tele health to India to increase the patient access to the health care information and to utilize the power of artificial intelligence needed to close the demand and provision of the human care providers. This chat app has led to lowering barriers to accessing health facilities and has gained far-sighted consultation to let go of timely cared and quality treatments, thus helping all the community more effectively.

Pharmabot: Pediatric Generic Medicine Consultant Chatbot

The paper introduces a pharma bot: Pediatric Generic Medicine consultant chatbot. It is a chatbot which is developed to prescribe, suggest and also give information on the generic medicines for children. The study introduces a computer application that act as medicine consults for the patients or parents who are confused with the generic medicines Science, Artificial Intelligence also called as machine learning and linguistics concerned with the interaction between computers and humans i.e. natural languages. In industries as well as academia, there is a need to understand and implement various language and computational linguistics knowledge so that it can be spread worldwide.

Pharmabot: Pediatric Generic Medicine Consultant Chatbot

The paper introduces a pharma bot: Pediatric Generic Medicine consultant chatbot. It is a chatbot which is developed to prescribe, suggest and also give information on the generic medicines for children. The study introduces a computer application that act as medicine consults for the patients or parents who are confused with the generic medicines. The researchers use Left and Right Parsing Algorithm to come up with the desired result. The researchers developed a medicine consultant chatbot, known as Pharmabot that will act as consultant pharmacist that will give the rational, appropriate and safe medication of generic drugs for children based on the information collected from the user by chatting. This system can be used by the parent of the patients who need medical assistance in taking the right generic medicine for certain ailment.

Automatized Medical Chatbot

Automatic medical interviews are built with discussion and technology in the mind by having the potential to reduce the costs of health care costs and also improve good access for medical and information services. We have created the diagnostic bots that involves patients in their clinical question and problem--solving issues based upon their findings and the profile. Our bot system is well suited for the identification of signals from the user input with an average accuracy of 65 percent. Using this obtained signal identified the appropriate symbols with a total memory of 65 percent and an accuracy of 71 percent. Finally, the bot has returned to the expected diagnosis of continued operation.

Chatbot for the Disease Prediction and Treatment Recommendation using Machine Learning

Hospitals are the most widely used when a patient receives clinical examinations, diagnosis and treatment recommendations. This also has become a practice for most of the people worldwide. People consider it a very reliable way to check them self's health status. The proposed idea is to develop an

alternative to the standard hospital visit and to create an appointment with a doctor for a diagnosis. This study aims to use the techniques of natural language processing and also the machine learning to create chatbot app. People who want medical assistance can contact chatbot as they contact with someone else and through the series of different questions, the chatbot will rectify the user's symptoms and thus, predicts a disease and recommend treatment. This program can be very helpful for people in doing daily tests, making people to know about their health status and encouraging people so that they take the right steps to stay healthy and fit. According to the research, such systems are less widely used and also less popular. Developing this framework can help many people avoid the time-consuming process of visiting the hospitals through this app for free, no matter where they live.

III. PROPOSED SYSTEM

The main objective of this work is to identify the patterns or features from the medical's data using the classifier model. The attributes which is similar to disease diagnosis can be observed. This will help the root causes of disease in depth.

The objective our works to predict the diagnosis of disease with number of attributes and provides solution to patient through chatbot. Here fourteen attributes involved in predicting disease. But fourteen attributes are reduced to six attributes by using three classifiers, Classification are used to predict the diagnosis of disease after the reduction of number of attributes.

Some chatbots are compact medical reference books which are useful not only for patients, doctors etc but also for those who want to learn something about health. The user feels that they are incorporated in the process of their health. Patients who feel included, who are interacting through chatbots with the healthcare system, will stay with the system, and that is important for them and the healthcare provider.

Some chatbots are compact medical reference books which are useful not only for patients, doctors etc. but also for those who want to learn something about health. The user feels that they are incorporated in the process of their health. Patients who feel included, who are interacting through chatbots with the healthcare system, will stay with the system, and that is important for them and the healthcare provider. The old chatbot are client communications systems and their best effort is a question-and-answer page on a website. Bot can facilitate to get the common health related question and prediction of disease without a human interference.

Natural Language Processing is subfield of a computational linguistics, artificial intelligence and Machine Learning. Since, computers play a great role in transmission and acquisition of information, there is a need to make computers understand natural languages. Technologies based on NLP are gaining widespread acceptance. e.g., Smart phones, other handheld devices are making use of translators, various machine learning approaches for retrieving text written in Chinese or Spanish. Language Processing is emerging to play a central role in this multilingual society. Natural Language Processing (NLP) is the field of Computer. Python has a wide range of standard libraries which makes it fit for performing computational and software engineering projects as well.

Python is a simple language and in this article we will be able to learn how a small and simple program helps in understanding and analyzing language data. How NLP concepts can be combined with Python in order to deduce the language concepts.

Natural Language Toolkit (NLTK)

NLTK originally created in the year 2001 as a part of a computational linguistics course in the Department of Computer and Information Science at the University of Pennsylvania. Since then this has been developed and then expanded with the help of many of contributors. It has now been adopted in courses in

many universities, and it serves as a basis of many research projects.

NumPy

NumPy is also Python package. Which stands for stands for the 'Numerical Python'. NumPy is a library consisting of multi-dimensional array objects and the collection of routines for processing of array. Numeric, the ancestors of NumPy, was developed by Jim Hugunin. Another package Numara was also developed, having some additional functionalities. In the year 2005, Travis Oliphant created a NumPy package by incorporating features of Numara into Numeric package.

Pandas

pandas are one of the Python packages that provides fast, flexible and descriptive data structures designed to make working with the "relationship" or "labeled" data easy. It aims to be a state-of the-art foundation for conducting effective, real-world data analysis in Python. In addition, it has the broad objective of being the most comprehensive and flexible data analysis / decryption tool available in any language. It is already on its way to this goal. pandas is well suited for many different kinds of data:

- Tabular data with the heterogeneously-typed columns, as in the SQL database table or Excel spreadsheet.
- Ordered and unordered (not necessarily fixed-frequency) time series of data.
- Arbitrary matrix data (homogeneously typed or heterogeneous) with the rows and columns labels.
- Any other form of observational / statistical data sets. Data actually does not need to be labeled at all to be placed in the pandas data structure These two main types of panda's data, Series (1-dimensional) and Data Frame (2-dimensional), handle most common cases of financial misuse, mathematics, science social, and many areas of engineering. For R users, Data Frame offers everything R's data. Frame offers and much more.

pandas is built on top of NumPy and aims to integrate well within the field of computer science and many other third-party libraries.

A Support Vector Machine (SVM)

SVM is the computer algorithm which learns by example to find the best function of classifier hyperplane to separate the two classes in the Input space. SVM is classified into two kinds of data, i.e. linearly and non- linearly separable data. Best hyperplane between two classes can be found by measuring the hyperplane margin and find out the maximum points

Naive bayes

Naïve bayes learning model applies Bayes rules. Every instance of data D is allotted to the class of highest subsequent probability. This model is trained through the Gaussian function with prior probability P Xf = priority $\in$ (0:
At the end, this testing data is classified based on the probability of association

Decision tree

For the training samples of data D, the trees are constructed on high entropy inputs. DAC trees are simple and fast constructed in a top-down recursive divide and conquer (DAC) approach. This tree pruning is performed to remove the irrelevant samples on D

User Login to System User registers on Chatbot application:

Then ask queries regarding to the health care and medical details.

A. **Ask some Questions:** You can ask some questions regarding some healthcare. And it's related to Text Input- text and text-Text Input conversation. Using Google API for inter conversion of text-Text Input and vice versa.

B. **Age based Medicine dosage:** details you can ask medical dosage related queries to this app in Text Input and system gets output for medicine API and speak out and display all data. Get your age from registration data and provide data related to your data like age, area, gender and so on. Give me age then predict disease using SVM Algorithm.

C. **Age based Medicine dosage:** details you can ask medical dosage related queries to this app in Text Input and system gets output for medicine API and speak out and display all data. Get your age from registration data and provide data related to your data like age, area, gender and so on. Give me age then predict disease using SVM Algorithm.

D. **Ask some Questions:** You can ask some questions regarding some healthcare. And its related to Text Input- text and text-Text Input conversation. Using Google API for inter conversion of text-Text Input and vice versa.

E. **Age based Medicine dosage:** details You can ask medical dosage related queries to this app in Text Input and system gets output for medicine API and speak out and display all data. Get your age from registration data and provide data related to your data like age, area, gender and so on. Give me age Then predict disease using SVM Algorithm.

F. **Get Medicine Details on medicine name:** You can ask about medicine related details on the basis of medicine names.

G. **Disease Prediction:** Depending on the disease symptoms SVM algorithm can predict the disease. Use Google API for Text Input-text and text Text Input conversion. The Chatbot API sends query to chatbot and get related answer and refer this answer analysis on that and display answer. Get medicine related data like medicine name, medicine expiry details and so on from medicine API. When user ask question to the scheme, logic of the complaint is recognized by applying LP.

Sense of the words is found using part of speech tagging and WordNet dictionary by using this sentiment analysis.

IV. IMPLEMENTATION

System design is the process of defining structure, components, modules, connectors, and system data to meet specific requirements. System design can be seen as applying theory to product development .Object-oriented analysis and methods become the most widely used methods in computer programming. System design is therefore a process of defining and developing systems to satisfy specific user needs. UML has become a common language in object analysis and focus.

Architectural design

System architecture is known as conceptual model which describes the structure and the behavior of the system. It comprises of different system components and the relationship describing how they work together to implement the whole system.
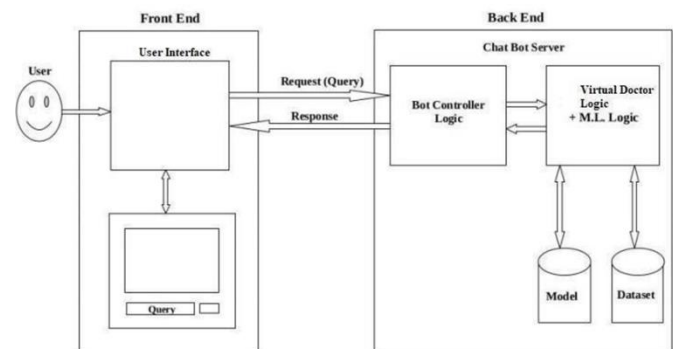
Fig 1: Architectural design is used to describes the structure and the behavior of the system.

Dataflow diagram

A data flow diagram is a clear demonstration of the "flow" of data through an information system, modeling its processes. DFD is often used as the first step tocreate an overview of the system without having to go into details, which can be explained later. DFDs can also be used for data processing. The DFD

indicates what type of information will be entered and exited from the system, how the data will proceed with the system, and where the data will be stored.

DFD level zero:

DFD-L0
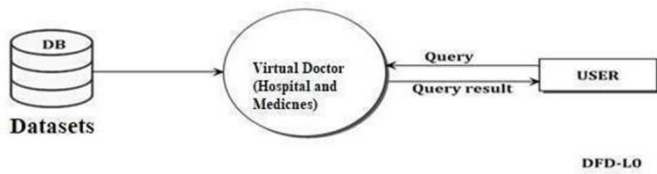
DFD level one:

Fig 2: Data Flow Diagram is used to describe the "flow" of data through an information system,
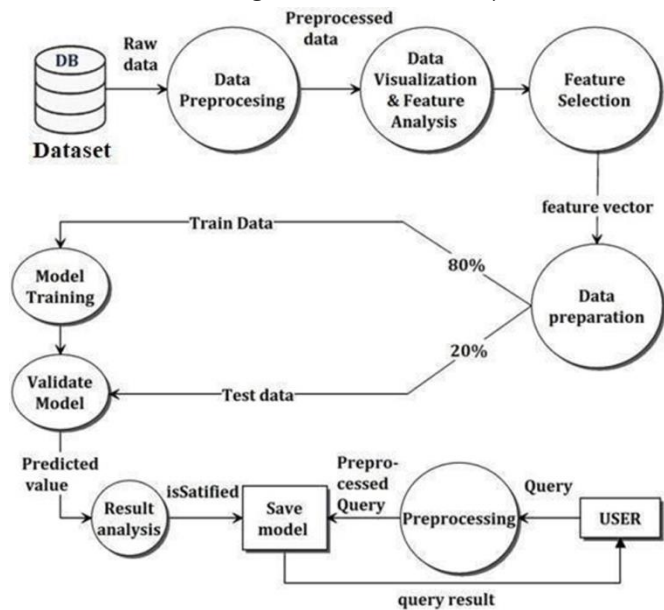
Use case diagram

A use case diagram at its simple is a representation of a user's interaction with the system which shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. While a use case itself might use case diagram can help provide a higher-level view f the system. It has been said before that "Use case diagrams are the blueprints for your system". They provide the simplified and graphical representation of what the system must actually do.
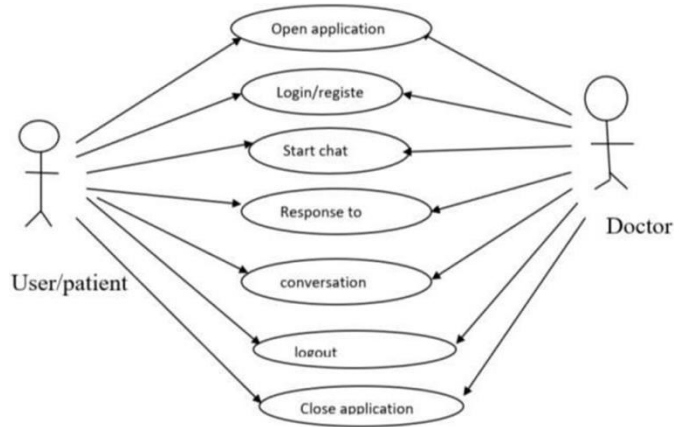
Fig 3: Use Case Diagram shows the relationship between the user and the different use cases in which the user is involved.

Sequence diagram
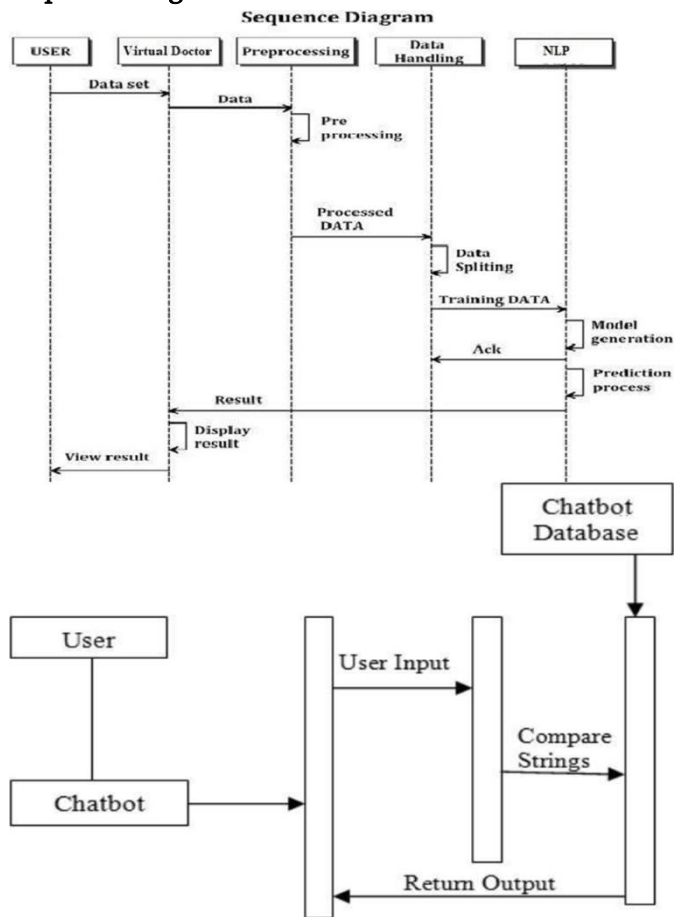
Fig 4: A sequence diagram shows object interactions arranged in a time sequence.

Activity diagram

Fig 5: Activity diagram describes the steps in use case diagram

Activity diagram visually represents the series of actions or flow of control in terms of chatbot conversation.
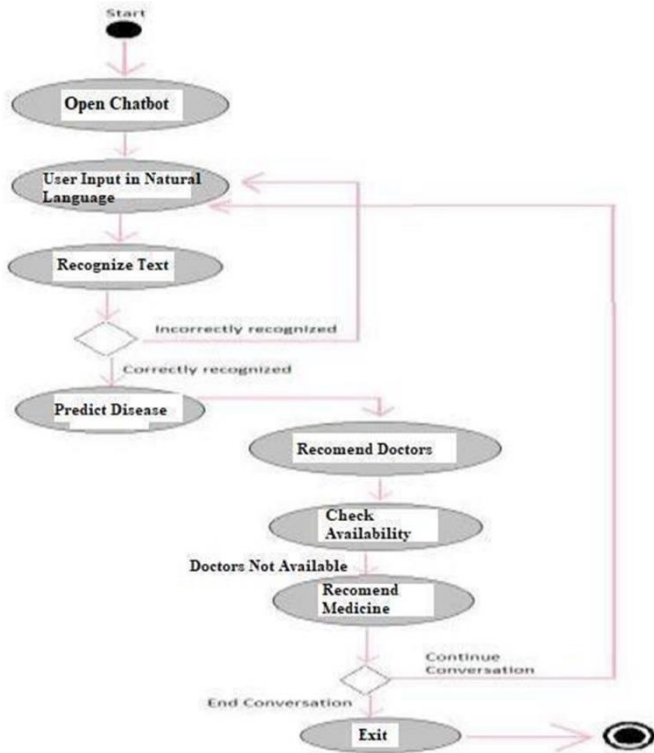
A. Front End

Web Development technologies like html, css and JavaScript is used. when we launch our project, we are directed to the web page where the user or the doctor options to be selected and then registration and login need to be done. After that in user section the chatbot will be there to which user can communicate if in emergency cases the user can directly communicate with the online doctors.

B. Back End

Python is used to support for back end and implement machine learning algorithms. Python is an interpreted language, which means you just type in plain text to an interpreter, and things happen. There is no compilation step, as in languages such as c or FORTRAN. To start up the Python interpreter, just type python from the command line on climate. You'll get a prompt, and can start typing in python

commands. Try typing in 2.5*3+5. and see what happens.

Fig.6: Decision Tree Implementation

Describes about the type of algorithm used. The first algorithm we have used is Decision Tree algorithm.

Fig7: Naive Bayes Algorithm Implementation.

The of second algorithm used in this project implementation is Naive bayes algorithm Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

Fig 8: Chatbot Implementation.

The Chatbot project uses Flask,

Flask is a web application framework written in Python. Where the smooth conversion is done between virtual doctor and the end user

```
import os
from flask import Flask, render_template, url_for,request, flash, redirect, session

from chatterbot import ChatBot
from chatterbot.trainers import ListTrainer

filenumber = int(os.listdir('saved_conversations')[-1])
filenumber = filenumber+1
file= open('saved_conversations/'+str(filenumber),"w+")
file.write('bot : Hi There! i am a medical chatbot. You can begin conversation by typing in a message and pressing enter.\n')
file.close()

app = Flask(__name__)

english_bot = ChatBot('Bot',
                storage_adapter='chatterbot.storage.SQLStorageAdapter',
                logic_adapters=[
                    {
                        'import_path': 'chatterbot.logic.BestMatch'
                    },

                ],
                trainer='chatterbot.trainers.ListTrainer')
english_bot.set_trainer(ListTrainer)
```

Fig 9: Doctor registration.

The initial step is registration of doctors with some details include username, email, specialists, address etc.

Fig 10: Doctor login

Once the doctor is registered then the doctor can login with the credentials given while registering

Fig 11: Doctor information and chat window

When the doctor enters to the dashboard by logging into application, his complete details is displayed with the chatting window linked below

Fig 12: User Login.

Now the 2nd type is user, the user has to register with his credentials and login into the application

Fig 13: User window and chatbot access.

Once the user is logged into the application then the list of doctors available online and offline is displayed for initiating the conversion

Fig 14: Chatbot before Conversation.

This page is displayed at the initial of the conversation with chatbot. Where user can chat with chatbot and discuss the type of illness.
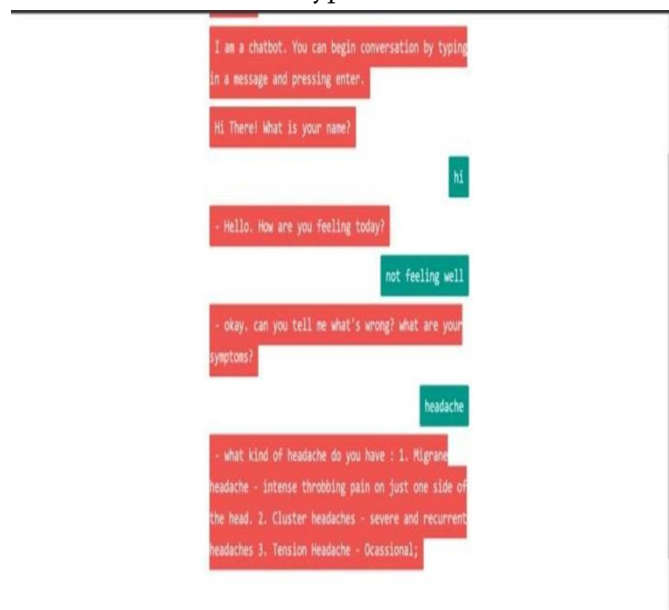
Fig 15: Chatbot conversation.

This page shows the conversation between the chatbot and the users and when the user discuss the symptoms the chatbot frequently responds with the solution
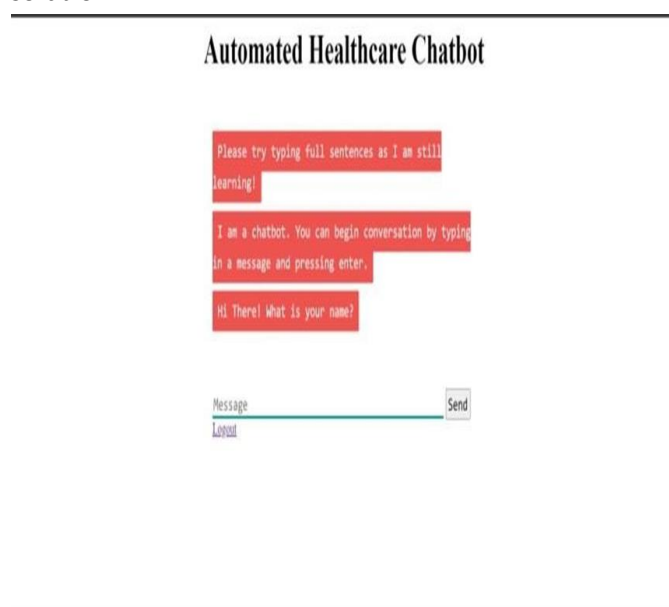
Tokenized text and pattern matching:

The main basic operations that can be applied for the text is tokenizing: breaking up with the stream of characters into words, punctuation marks, numbers and another discrete item. So, for example the character string "Dr. Watson, Mr. Sherlock Holmes", said Stamford, introducing us. can be tokenized as in the following example, where each token is enclosed in single quotation marks: '"' 'Dr.' 'Watson' ',' 'Mr.' 'Sherlock' 'Holmes' '"' ',' 'said' 'Stamford' ',' 'introducing' 'us' '.' In this level, words have not been classified into grammatical categories and we have little indication of syntactic structure. Still, a fair amount of information may be obtained from relatively shallow analysis of tokenized text. For example, suppose we want to develop a procedure for finding all personal names in a given text. We know that personal names always start with capital letters, but that is not enough to distinguish them from names of countries, cities, companies, racehorses and so on, or from capitalization at the start of a sentence.

Parts of speech:

A number of different POS classifications have been developed within computational linguistics and we will see some examples in subsequent chapters. The following is a list of categories that are often encountered in general linguistics: you will be familiar with many of them already from learning the grammar of English or other languages, though some terms such as Determiner or Conjunction may be new to you.

Noun fish, book, house, pen, procrastination, language Proper noun John, France, Barack, Goldsmiths, Python Verb loves, hates, studies, sleeps, thinks, is, has Adjective grumpy, sleepy, happy, bashful Adverb slowly, quickly, now, here, there Pronoun I, you, he, she, we, us, it, they Preposition in, on, at, by, around, with, without Conjunction and, but, or, unless Determiner the, a, an, some, many, few make the standard distinction that nouns 'generally refer to people, places, things or concepts' while verbs 'describe events or actions'. This may be helpful when one is starting to learn grammatical terminology but is something of an over-simplification. One can easily find or construct examples where the same concept

can be expressed by a noun or a verb, or by an adjective or an adverb. And on the other hand, there are many words that can take different parts of speech depending on what and how they can do in a sentence:

1. Rome fell swiftly.
2. The fall of Rome was swift.
3. The enemy completely destroyed the city.
4. The enemy's destruction of the city was complete.
5. John likes to fish on the river bank.
6. John caught a fish.

Additionally, some types of verbs do not correspond to any particular action but serve a purely grammatical function: these include the auxiliary verbs such as did, shall and so on. So, in summary, we can often only assign a part of speech to a word depending on its function in context rather than how it relates to real things or events in the world.

V. CONCLUSION

This project describes a clinical discussion that can be used to replace all the standard diagnostic method and treatment recommendations. Chatbot work as a virtual doctor. Chatbot also works as a user application. In this application the user can specify his or her disease symptoms in the chatbot and then the chatbot can tell the health measures to be taken next. General information about the symptoms and diseases are available in the database so the chatbot example can provide information about his/her diseases and treatments to the user. After recognizing the symptoms of different users, chatbot finally predict the disease of the user and provide a link where the clinical information is visible. A smart medical chatbot can be useful for the patients by identifying their symptoms as they are explained, providing the proper diagnosis and providing correct treatment for the disease.

VI. REFERENCES

[1]. Bhagya Shree S R, Chandra Shekar P, 'Automated Medication Dispensing System', 2015 IEEE.

[2]. KahtanAziz, 'Real-Time Healthcare Monitoring and Tracking System using GSM/GPS Technologies', August 2016 IEEE.

[3]. Mahaveer Penna, 'Design and Implementation of Automatic Medicine Dispensing machine', May 2017 IEEE.

[4]. Vishal Tank, Sushmita Warier, 'Medicine Dispensing Machine Using Raspberry Pi and Arduino Controller ', March 2017 IEEE.

[5]. Petkus, Vytautas, AlfonsasVainoras, Kristina Berskiene, ZenonasNavickas, RimtautasRuseckas, Mantas Deimantavicius, and ArminasRagauskas. "Method for Prediction of Acute Hypotensive Episodes." ElektronikairElektrotechnika 22, no. 1 (2016): 44-48.

[6]. Bhattacharya, Satyajit, Vaibhav Ranjan, and Vijay Huddar. "A novel classification method for predicting acute hypotensive episodes in critical care." In Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 43-52. ACM, 2014.

[7]. Ghassemi, Marzyeh. "Methods and models for acute hypotensive episode prediction." PhD diss., MSc Thesis, 2011.

[8]. Kim, Sun-Hee, Lei Li, Christos Faloutsos, Hyung-Jeong Yang, and Seong-Whan Lee. "HeartCast: Predicting acute hypotensive episodes in intensive care units." Statistical Methodology 33 (2016): 1-13.

[9]. Ghosh, Shameek, Mengling Feng, Hung Nguyen, and Jinyan Li. "Hypotension risk prediction via sequential contrast patterns of ICU blood pressure." IEEE journal of biomedical and health informatics 20, no. 5 (2016): 1416-1426.

Feature Selection using Nature Inspired Modified Firefly Algorithm

Shashikala B[1], R. Saravanakumar[2]

[1]Department of CSE, B.T.L Institute of Technology & Management, Bengaluru, Karnataka, India
[2]Department of CSE, Dayananda Sagar Academy of Technology & Management, Bengaluru, Karnataka, India

ABSTRACT

Feature Selection, a feature engineering process for identifying the relevant attributes of the given dataset with the aim to reduce the curse of dimensionality especially when handling big data sets that contain numerous numbers of attributes. There are various methods to find the optimal number of attributes. The motivation behind this work is to develop a efficient method for finding the pertinent attributes by using the modified firefly algorithm thereby improving the efficiency of the classification task. The Firefly Algorithm (FFA) is a nature inspired swarm intelligence meta-heuristic algorithm, developed based on the flashing behavior of fireflies during their journey. This paper work involves the variant of the basic algorithm where partitioning the population is employed with independent processing each of the sub-population. The first phase involves partitioning, processing and finding the optimal set of attributes from each sub population and in next phase involves integrating the best results of each subpopulation and evaluating the performance. The output is substantiated with the measures like f percentage of accuracy and count of selected attributes. Three datasets were used for the experiment. The experimental results conducted on public real dataset demonstrate that the modified firefly algorithm show a significant improvement in selecting optimal set of attributes.

Keywords—Feature Selection, Firefly, Accuracy, Classification, Partitioning

I. INTRODUCTION

These days, various disciplines end up using larger dimensional data because of the advent of various technologies like IoT, AI , etc. , hence extracting knowledge becomes a challenging task for most machine learning algorithms. Feature selection is the method that aims at identifying relevant and removing unnecessary features which may pose a threat to the classification   performance.One major reason is that machine learning follows the rule of   "garbage in-garbage out" and that is the reason for giving more importance to the data that is being fed to the model.

Selecting the right set of features to be used for data modeling has been shown to improve the performance of supervised and unsupervised learning.

However, conventional methods lag in scalability while handling larger datasets and to derive useful results. Feature selection techniques are normally when there exists many characteristics with less instances which may help in getting less complex machine learning for better understanding within for shorter training times and to get rid of the curse of higher dimensions.

The principle behind retaining the relevant features and removing others is that , it will not result in loss of greater information.

The swarm optimization algorithms having simpler processing with proficient searching strategy have become popular in solving many performance focused optimization problems. To name a few swarm intelligent algorithms like Particle swarm optimization (PSO) , Ant Colony optimization (ACO) , Artificial Bee Colony algorithm (ABC) , Simulated Annealing (SA) algorithm , Bacterial Colony Chemo taxis (BCC) , FireFly Algorithm (FA) for optimization. This work presents a methodology for identifying the relevant features using the Firefly algorithm for reducing the data dimension.

II. RELATED WORK

Many researchers still work on feature engineering tasks like extraction, selection, etc., wherein the previous works helps to draw some useful inferences. Xiao-Yu Zhang et al. [1] proposed an ensemble algorithm for Malware Detection which extracts the best features from high-dimensional features that represent the data. The inference is that feature selection is used along with other data mining task to reduce the computational burden.

Ahmad Riza'ain Yusof et. Al[2] developed a efficient model for Intrusion Detection with the help of feature selection along with other machine learning algorithms . The model combined two elements namely Consistency Subset Evaluation (CSE) and DDoS Characteristic Features (DCF). The outcomes obtained proved the efficiency of the proposed model. This work draws the inference that feature selection can be applied for various types of data.

Shruti Gupta et. [3] worked on SAR , land cover classification task by using class wise feature subset selection .The author performed the classification using the separability index and Naive Bayes classifier approach. The class wise feature selection improved in a better way in terms of accuracy when compared with commonly used techniques with a lesser computational cost.

Xue Zhang, Zhiguo [4] used a hybrid model by clubbing one-class F-Score method, improved F-Score method and genetic algorithm for identifying the relevant features for handling unbalanced data seeking the optimal classification model. It was found that the time efficiency of the work was better compared to conventional method but searched for the best feature combination in larger feature space. Thus this work gives an idea of using evolutionary method for feature selection.

Nur Farahlina Johari et.al.[6], discussed how swarm intelligent Firefly Algorithm (FA) can be applied to solve various optimization problem such as discrete, chaotic natures and multi-objective type.

E. Emary et.al in the paper " Firefly Optimization Algorithm for Feature Selection" [7] proved FFA being robust and stable in comparison with GA and PSO.

Vandana Agarwal and Surekha Bhanot [8] had analyzed the merits of the algorithm as the fast convergence with better performance. Thus this research work results give the scope of FA for feature selection.

D. Sheela Jeyarani, A. Pethalakshmi[9] had proposed a ensemble model using fuzzy entropy with firefly concept to select quality features. The algorithm's performances were analyzed using four different high dimensional data sets WILT, ORL, LC and LTG and the results shown prove that the algorithm out performs the traditional way.

Anbu , G. S. Anandha Mala [10] developed a model for software defect prediction using swarm intelligent firefly algorithm and evaluated against classifiers like SVM, NB and KNN. Results were promising with FA . Huali Xu , Shuhao Yu , Jiajun Chen , Xukun Zuo [11], had put forward an improved Firefly Algorithm using dicsrete coding for performing feature election for classification in wrapper mode . The improved FA employs opposition-based learning in population initialization and opposition strategy in the searching

process which had fastened the convergence rate to obtain the global optima.

Li Zhang, Kamlesh Mistry, Chee Peng Lim, Siew Chin Neo [12] had proposed a modified FA model which had shown superior discriminative capabilities and outperformed FA variants, classical methods for diverse feature selection problems. The statistical test results ascertain the effectiveness of the proposed FA variant in enhancing classification and regression models for supporting decision making processes.

III. FEATURE SELECTION

There are many models for performing the task of identifying minimum number of relevant features from a larger feature space, thereby reducing the dimensionality of data. This task is performed with intention of getting better results in various data mining techniques such that it is as close as possible (or even better) when obtained using all attributes of data[5]. This dimensionality reduction will enhance the understandability of the extracted pattern thereby making the posterior learning much faster.

There are four different categories of methods which can be used for performing feature selection.

(i) **Wrapper method**: In wrapper method the selection criterion is part of the objective function and relies on the machine learning algorithm that is used to estimate the precision of the selected features.

(ii) **Filtering method**: The selection is based on data related measures, such as separability or crowding. A subset of features is selected by this approach without using any learning algorithm. Higher-dimensional datasets use this method and it is relatively faster than the wrapper-based approaches.

(iii) **Embedded method**: In embedded method, the selection of subset of relevant features is performed during the classifier model construction. The learning algorithms determine the specificity of this approach and it selects the features during the process of training the data set.

(iv) **Hybrid approach:** This approach combines filter and wrapper-based methods to take the advantages of both the categories. This approach first selects the possible optimal feature set which is further tested by the wrapper approach.

Some of the authors developed an innovative way to select relevant attributes in large dimension dataset and designed to be deployed in a single machine but was not scalable.

Few researchers in order to handle the drawback of conventional methods employed evolutionary algorithms to perform the same. In such approach usually, the set of features were encoded in the form of binary vector, where presence of one or zero at each position determines if a respective feature is selected or not. This kind of representation helps to achieve exploration and exploitation nature of evolutionary algorithms.

IV. FIREFLY ALGORITHM

Firefly Algorithm (FA) is a meta-heuristic nature inspired algorithm developed based on the interesting flashing behavior of fireflies during the movemet . Meta heuristic methods are normally used for optimization problems which are NP hard.

It mainly works on three principles :

- Fireflies irrespective of gender attracts other fireflies based on their attraction power
- The firefly can attract and guide other firefly by the brightness of their flashing light which is inversely proportional to distance.
  If there exists no flashing light to guide then firefly will move randomly in any direction.
- The fitness function determines the light intensity which in turn gives brightness of a firefly.

In this method, each firefly represent an optimal set of features and will possess a brightness value which is the value obtained by the fitness function which

evaluates the firefly content. New solution evolves as the fireflies move towards the brighter firefly.

During the evolution process, better fireflies are retained and carried over to the next successive step. The fireflies are graded based on the objective function output.

The distance between the fireflies are found by using the Euclidean distance formula and the content of firefly can be updated based on the Intensity and light absorption co-efficient.

Suppose that for each solution i , i(Xi) represents the position in solution space at that point of time w.r.t iteration. If the fitness of ith solution is greater than another solution j , the space between them in the solution space can be found by using Euclidean distance formula as mentioned in Equation (1).

r_{ji}= sqrt(for each position $(X_i - X_j)^2$)………..….(1)

The value of Equation (1) will be used in Equation (2) to calculate the new attractiveness value.

$\beta = \beta_0\, e^{-\gamma r_{ij}^2}$ ....(2)

Where β_0 represent the attractiveness when fireflies are at same position and normally set to 1. γ is the light absorption coefficient and normally set to 1 and rand represent the random number.

The new location of ith firefly by using the below equations(3) and (4)

$NewX_{ij}$ = Old X_i + β.rand.ΔX_{ij} + α (rand -0.5)..........(3)

ΔX_{ij} = $(X_j - X_j)$ …………………..………………........(4)

Thus in FA the evolution of solutions in terms of fireflies will repeat until the specified the termination criteria.

V. FEATURE SELECTION USING FIREFLY

This section explains how firefly algorithm can be used to the select the optimal set of features for a given dataset.

A. Methodology

This research work involves two phase as shown in Fig.1. wherein the first phase focuses on firefly algorithm to select the optimal features and the second phase focuses on evaluating the accuracy of firefly algorithm.
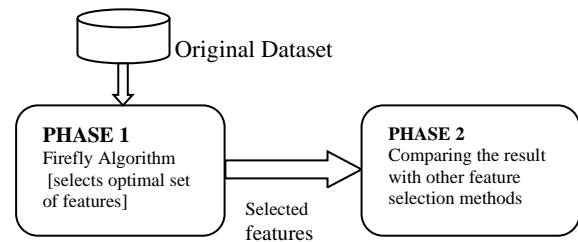
Fig.1. Outline of Proposed Methodology

In the Firefly Implementation the size of the firefly is decided based on the original number of attributes of the given dataset. In this work the firefly is binary encoded where 1 indicates the selection of attribute and 0 indicates the absence of attribute which is shown in the following Fig.2

| 1 | 0 | ……………… | 1 | 1 |

Fig.2. Representation of a Firefly

Based on the number and position of 1's in each firefly the dataset will be reduced and the brightness of each firefly will be the calculated by the objective function.

The objective function will output the accuracy of the classifier which takes the selected attributes w.r.t each firefly and applies to the input dataset to reduce the dimension and generates a classifier based on that. The performance of proposed method is validated against the results of other conventional feature selection methods by taking the input (that is the count of selected features) from phase1.

The experimental results show that firefly algorithm select better set of features compared to other methods.

B. Algorithm

The Phase1 of the proposed methodology involves the following algorithm

Firefly Algorithm For Feature Selection
_____

Input: Dataset
Output: Optimal Set of selected features
_____

Initialization: Attractiveness at zero distance β0=1
      Light absorption coefficient   γ=1
      Objective Function f(x): x=<$x_1,x_2,..x_d$>
[d represents number of attributes of dataset]
        Max_Iterations=<Integer constant value>
  mber of fireflies: n=<Integer constant value>

Step: 1 Create a pool of " n" fireflies
Step: 2 evaluate the light Intensity of fireflies using f(x)
Step: 3 Loop till Max_Iterations
Step:3.a.Loop for i=1 till n
          Loop for j=1 till n
              If(Intensity(j) > Intensity(i))
Step :3b.   Move i[th] firefly towards jth using Eq.(3)
Step :3c   Vary the attractiveness using Eq.(2)
Step :3d   Evaluate the fireflies and update light intensity
Step :3e   Select and store the best firefly
Step:4   Output the best firefly
Step :5   Count the number of selected features

==
The objective function used in the algorithm is detailed below in Fig.3.

Objective function: f(x)
_____

Input: A Firefly of size'd'
Output: Accuracy of classifier
_____

Step1: Scanning the firefly to select those attributes which are set by the corresponding position in the firefly
Step2: Based on selected attributes reducing the dataset dimension

Step3: Generating a decision tree classifier model using reduced dataset
Step4: Calculating the accuracy of classifier

==
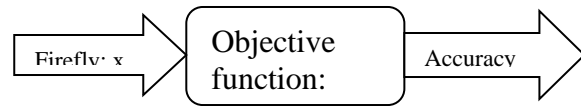
Fig.3. Objective function Outline

The output of the best firefly is considered for the second

phase Wherein the number of features to be selected is calculated based on the firefly and fed as input to other feature selection methods like Principal Component Analysis , Selectkbest , Univariate Methods .Then   the result of firefly accuracy and other feature selection methods are compared .

C. Experimental Results
The proposed methodology is implemented using python language in jupyter notebook. Three dataset has been used which has the characteristics as mentioned in Table I.
In this work the number of fireflies was assumed to be 10 with maximum iteration value as 20. The accuracy of the algorithm is expressed in the Table II.

TABLE I
DATASET DESCRIPTION

S.No	Experimented Dataset		
	Dataset Name	Instances	Attributes
1	MobilePrice.csv	1400	22
2	Winequality-white.csv	4898	12
3	WDBC.csv	569	32

TABLE II

FEATURE SELECTION IN MOBILEPRICE.CSV

Methods	No. of features:21+1(target)=22		
	Features	Accuracy %	Selected Features / Variance
DT	All	80.47	All the features are selected
PCA	9 PC	84.04	[6.15630930e-01 1.50304460e-01 1.02199334e-01 8.21454322e-02 4.88303712e-02 6.52398618e-04 1.70246904e-04 2.39451178e-05 1.59354488e-05]
Select K Best	10/22	82.38	[0 1 7 9 12 13 14 16 17 21]
FFA	10/22	84.52	[1, 2, 3, 6, 8, 12, 13, 14, 19, 21]

TABLE III

FEATURE SELECTION IN WINEQUALITY-WHITE.CSV

Methods	No. of features:11+1(target)=12		
	Features	Accuracy %	Features
DT	All	59.25	

Methods	No. of features:11+1(target)=12		
	Features	Accuracy %	Features
			All the features are selected
PCA	4 PC	58.57	[9.11312392e-01 7.73858874e-02 1.04363597e-02 5.13705969e-04]
Select K Best	5/12	59.45	[3 5 6 10 11]
FFA	5/12	61.9	[1, 8, 9, 10, 11]

TABLE IV

FEATURE SELECTION IN WDBC.CSV

Methods	No. of features:31+1(target)=32		
	Features	Accuracy %	Features
DT	All	90.05	All the features are selected
PCA	20 PC	92.39	[1.00000000e+00 2.47308925e-11 4.42650282e-13 4.73483824e-14 2.93772913e-15 2.23365907e-15 1.86190927e-16 1.02637986e-16 2.26307710e-17 9.56742320e-18 4.13516023e-18

Methods	No. of features:31+1(target)=32		
	Features	Accuracy %	Features
			1.56967553e-18 3.82001569e-19 1.50312367e-19 1.01095990e-19 7.90576666e-20 3.58412249e-20 1.70676362e-20 1.17907981e-20 9.74524449e-21]
Select K Best	21/32	91.22	[0 1 2 3 4 6 7 8 11 13 14 17 21 22 23 24 26 27 28 29 31]
FFA	21/32	97.66	[0, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 20, 22, 24, 25, 27, 31]

The last column of Table II , III and IV represents the selected features or the variance with respect to specified method in first column. In all these tables the number of fireflies considered is 10. In the last column the number specified indicates the selection of attribute represented by the corresponding index of the array whereas for PCA method the numerical values indicate the variance of the principal components.

VI. CONCLUSIONS

Thus from the experimental results it is evident that for NP hard optimization problem like feature selection, the Meta heuristic firefly algorithm performs better compared to other deterministic methods. The drawback that is observed during experimentation is that, to get a consistent output and to explore and exploit the solution space the algorithm has to incorporate a larger number of iteration and fireflies which will take higher execution time. Thus the drawbacks pave a way for further research.

VII. REFERENCES

[1]. Xiao-Yu Zhang , Shupeng Wang, Lei Zhang,Ensemble Feature Selection with Discriminative and Representative Properties for Malware Detection,IEEE Infocom 2016 Poster Presentation

[2]. Ahmad Riza'ain Yusof ,Hazlina Hamdan,Nur Izura Udzir,Mohd Taufik Abdullah ,Ali Selamat, Adaptive Feature Selection for Denial of Services (DoS) Attack , 2017 IEEE Conference on Application, Information and Network Security (AINS)

[3]. Shruti Gupta, Sandeep Kumar, Akanksha Garg, Dharmendra Singh, N S Rajput Indian Institute of Technology Roorkee, Roorkee (UK), India , CLASS WISE OPTIMAL FEATURE SELECTION FOR LAND COVER CLASSIFICATION USING SAR DATA, 978-1-5090-3332-4/16/$31.00 ©2016 IEEE

[4]. Xue Zhang, Zhiguo Shi*, Xuan Liu, Xueni Li , A Hybrid Feature Selection Algorithm For Classification Unbalanced Data Processsing, , 978-1-5386-8543-3/18/$31.00 ©2018 IEEE DOI 10.1109/SmartIoT.2018.00030

[5]. H. Liu and H. Motoda, Computational Methods of Feature Selection, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Chapman & Hall/CRC Press, 2007.

[6]. Nur Farahlina Johari, Azlan Mohd Zain, Noorfa Haszlinna Mustaffa, Amirmudin Udin, Firefly Algorithm for Optimization Problem, Applied Mechanics and Materials Vol. 421 (2013) pp 512-517© (2013) Trans Tech Publications, Switzerland

[7]. E. Emary et.al , Firefly Optimization Algorithm for Feature Selection , ACM New York, NY, USA ©2015 , ISBN: 978-1-4503-3335-1 doi>10.1145/2801081.2801091

[8]. Vandana Agarwal and Surekha Bhanot ,Firefly Inspired Feature Selection for Face Recognition ,978-1-4673-7948-9/15/$31.00 ©2015 IEEE

[9]. D. Sheela Jeyarani, A. Pethalakshmi,Optimized Feature Selection Algorithm for High Dimensional Data, Indian Journal of Science and Technology, Vol 9(26), DOI: 10.17485/ijst/2016/v9i31/79656, August 2016

[10]. M. Anbu , G. S. Anandha Mala , Feature selection using firefly algorithm in software defect Prediction , © Springer Science+Business Media, LLC 2017

[11]. Huali Xu , Shuhao Yu , · Jiajun Chen , Xukun Zuo,An ,Improved Firefly Algorithm for Feature Selection in Classification , Wireless Pers Communication, China , Springer Science+Business Media, LLC, part of Springer Nature 2018

[12]. Li Zhang, Kamlesh Mistry, Chee Peng Lim, Siew Chin Neoh , Feature Selection Using Firefly Optimization for Classification and Regression Models, ELSEVIER, Decision Support Systems,Volume 106, February 2018, Pages 64-85, https://doi.org/10.1016/j.dss.2017.12.001 Pan, W.T., A New Fruit Fly Optimization Algorithm: Taking the Financial Distress Model as an Example. Knowledge-Based Systems, 26, 69 – 74 ,2011
http://dx.doi.org/10.1016/j.knosys.2011.07.001

Shillelagh – The Smart Stick

Sriram S[1], Roshini S[1], Premalatha D[1], Deepti Rai[1]

[1]Department of Computer Science Engineering Department, New Horizon College of Engineering

Bengaluru, Karnataka, India

ABSTRACT

The visually impaired individuals face numerous challenges in their normal life. Eventually, they rely on others for their needs, which makes them hesitant in a new environment. Even in E-Commerce websites and numerous business vendors sell the so-called smart stick, which are limited to basic functionality and are void of any advanced technology like navigation, object identification, object detection, etc. Hence with absence of these any modern feature it makes them heavily dependent on another individual or their memory even for performing everyday chores. Hence by integrating the latest technologies like object detection, navigation, obstacle detection, etc we aid the visually challenged people to interact with the environment better than ever before. With consideration of all the above-mentioned points we have created our project Shillelagh – The Smart Stick which focuses to resolve the problem faced by the blind and visually impaired people. It integrates numerous cutting-edge technologies to develop a modern solution which can make their life simpler and much better. The project uses the concept of Navigation, Object recognition, object detection, depth sensing, all of these seamlessly integrated under one mobile application which is specially designed for the blind people to interact with the world.

Keywords — Navigation, Obstacle Detection, Object Identification, Visually Impaired, Smart stick

I. INTRODUCTION

In our world the people are diverse, each one has their own strength and weakness among them few people are physically challenged, these people shouldn't hinder their growth and opportunities. These people are one among us who inherit a great talent and calibre and it is our duty to give them an equal opportunity and ensure these people too have their own independent life. We must address their problem such as, performing daily chore as These individuals need to rely on their memory to find their possessions and may get disturbed in the event that somebody misplaced any item or it tumbles down accidentally. It becomes increasingly difficult for them to solve these problems without the visual perception.

Moreover, even in E-Commerce websites numerous business vendors are selling the so-called Smart walking-stick or the sticks for the visually impaired individuals which are restricted to minimal functionalities and does not include any advanced feature like navigation, object recognition, object detecting, and so on. Even with the rapid technology growth there has been very limited innovation for these sticks and there's not been any improvement in the smart stick or a stick which can really help the visually challenged person more independent.

Thus, our Project "Shillelagh – The Smart Stick" which centres to aid the visually impaired and coordinate various cutting-edge technologies to make their life independent and much better. The project utilizes the concepts of object identification, navigation, obstacle detection, etc and integrate all these as a single unit controlled by a mobile application developed specially for the visually impaired to interact with the world better and be independent.

II. RELATED WORKS

In providing the right equipment for blind, Jismi Johnson and team in paper [1] have told that It allows the user to walk freely by detecting obstacles in front of him. Images will be captured using a camera and the camera is connected to the Raspberry Pi. If any obstacle comes in front of blind person, he can know about the obstacle

To assist blind people, The Intelligent cane designed to help blind people avoid obstacles by using onboard sensors, Gurubaran, and team in paper [2] have proposed an idea of designing electronic stick using Global System Messaging (GSM), Global Positioning System (GPS) and Ultra-sonic technology.

To classify obstacles ahead using ultrasound sensor, Deepika S and team in paper [3] have designed a cane to detect obstacle for collision avoidance, it also detects for the object in directions up, down and front. The other sensor placed near bottom tip of the walking cane to find the pits on the ground.

In delivering a smart cane that helps people with visual challenges to safely navigate with advanced technology, Kher Chaitrali S and team in paper [4] have proposed a navigation device for the visually impaired which is focused on providing voice output for obstacle prevention and navigation using infrared sensors, RFID technology, and android devices.

With the intent to enhance visual navigation for visually disabled individuals, Shubham Adhe and team in paper [5] have designed the blind walking stick that is incorporated with addition to light and water sensing with an ultrasonic sensor. The ultrasonic sensors used to detect obstacles with the ultrasonic waves were the initial use of our proposal. In this project they are using ultrasonic sensors, water sensor, buzzer, Bluetooth module.

An android application built to read and update the location using RFID tags, Sukriti Sudhakar and team in paper [6] have designed the smart cane that has sensors embedded to it, thereby it senses the objects/intruder, when any objects or obstacles come in range of an ultrasonic sensor then the person is alerted with a quick response time.

III. METHODOLOGY

The main features of this Application are:
1. Navigation
2. Obstacle Detection
3. Object Identification

- Firstly, the user powers on the Shillelagh-The smart stick device and it gets connected with the android application.
- A system check is performed to ensure that all the components are in working state.
- If they are not working then a voice message is sent to the user to rectify the error.
- If all the components are working, then the sensors get activated and detects for any obstacle.
- If depth ahead it detects using the Ultrasonic sensor and basically alerts the users about it.
- The system waits for user request for object identification or navigation or to power off the Shillelagh-The Smart Stick.
- In case if the user requests for object identification, first the camera embedded on the stick is turned on and it starts sending live feed data to the application for further processing for object detection and identification
- After the application has identified the object, the object name will be conveyed to the user.

- This process will continue until the user asks to stop.
- If the user request for navigation, the current location will be automatically extracted through the GPS and destination will be obtained from the user.
- The navigation function in the application will be enabled and proper direction commands or instructions will be providing to the user which is voice enabled till he reaches the destination.
- At any point in time if the user wants to switch off the walking stick, they can do so by requesting for power off, then the Shillelagh-The Smart Stick will stop all the ongoing process and prepare for power off.

IV. DESIGN

Fig 1: Flowchart of Shillelagh-The Smart Stick

Fig 2: DFD of Shillelagh-The Smart Stick

Firstly, as soon as the user powers on the device, the sensor gets activated and detects for any obstacle or depth ahead using ultrasonic sensor. Then upon request for navigation which is invoked through mobile application, guides for the direction for the given destination based on Google Api that is voice enabled. Along the path the user will be conveyed about the things around him for object identification through machine learning framework. It detects for objects in real time from camera feed within image frame.

V. IMPLEMENTATION

The user has to sign in to the mobile application by providing email address and password. The dashboard will be enumerated with three options – Navigation, Object detection and Emergency. If the user wants to select Navigation, the user has to single tap on the screen and then double tap to confirm their selection. The user will be asked for destination following which the application performs speech recognition to identify the destination and gives proper direction commands with the help of google maps. In case of emergency, it calls the emergency number given during registration. For Object detection, the camera embedded on the stick will power on and send live feed data to the application. Then the application will process the data and convey the object name to the user.

The stick is also equipped with sensors for obstacle detection. It uses ultrasonic waves to determine the distance between the stick and ground. Hence, whenever stairs or pothole or obstacle is ahead of them, there is a sharp increase in the distance and the sensor alerts the user with a buzzer and vibration.
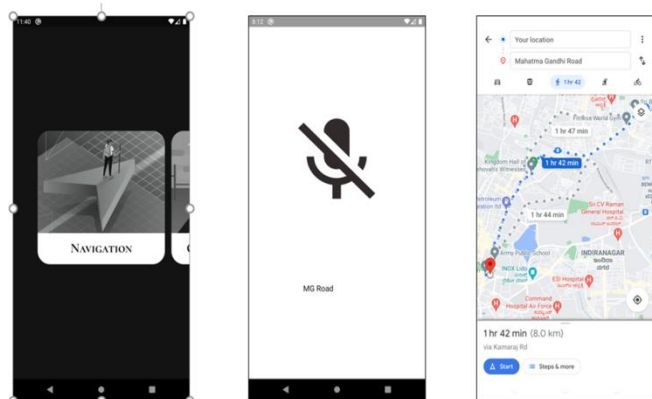
Fig 3: Output of mobile application
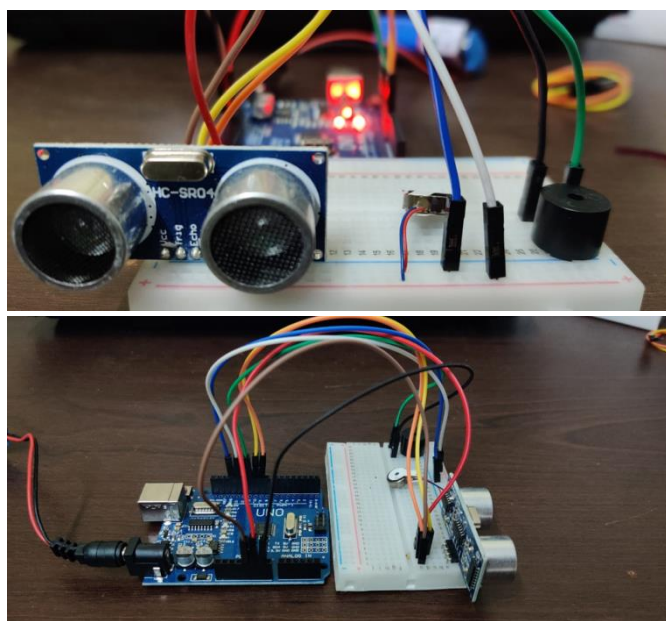
Fig 4: Hardware components

VI. CONCLUSION

To enhance independence of the visually challenged people. By integrating the latest technologies for object detection, navigation, obstacle detection, etc to enable these visually challenged people to interact with the environment better than ever before. Hence, we have created our project Shillelagh – The Smart Stick focuses to resolve the problem faced by the blind and visually impaired people and integrate numerous cutting-edge technologies to develop a modern solution which can make their life simpler and much better. The project uses the concept of Navigation, Object recognition, object detection, depth sensing, all of these seamlessly integrated under one mobile application specially designed for the blind people to interact with the world better than ever before.

VII.   FUTURE SCOPE

1. We can deploy the entire application into a smart watch instead of a mobile device.
2. We can integrate google assistant or any other virtual assistant with AI capabilities.
3. Add a fingerprint encryption to access the application.
4. The Braille input device gives the blind person an uncomplicated method to provide the destination address for navigation.
5. We can integrate it to IOS platform
6. More sensors can be added to enhance the features of the application.
7. For object identification, we can use the raspberry pi camera for sending live-feed data to application.

VIII. REFERENCES

[1]. Jismi Johnson, Nikhal Rajan P, Nivya M Thomas, Rakendh C S and Sijo T Varghese, "Smart Walking Stick for Blind", International Journal of Engineering Science Invention R&D Vol. III, Issue IX, March 2017.
[2]. Gurubaran , Gowrishankar Kasilingam, Mritha Ramalingam, "A Survey of Voice Aided Electronic Stick for Visually Impaired People", International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 8 (September 2014) .

[3]. Deepika S, Divya B.E, Harshitha K, Komala B.K, Shruthi P, "Ultrasonic Blind Walking Stick", International Journal of Advance Electrical and Electronics Engineering (IJAEEE) Volume-5 Issue-6 2016.

[4]. Sukriti Sudhakar, "Smart Cane for Visually Impaired", International Journal of Engineering Science and Computing, August 2018.

[5]. Shubham Adhe1 ,Sachin Kunthewad2 ,Preetam Shinde3 V.S.Kulkarni4, "Ultrasonic Smart Stick for Visually Impaired People", IOSR Journal of Electronics and Communication Engineering

[6]. Kher Chaitrali S., Dabhade Yogita A., Kadam Snehal K., Dhamdhere Swati D., Deshpande Aarti V, "An Intelligent Walking Stick for the Blind ", International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.

Agro Vision - Crop Yield Prediction and Crop Leaf Disease Detection

Payal Mankotia[1], Sankeerthana M[1], Siri Soundaraya S A[1], Mr. K.Kiran Kumar[2]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Senior Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Machine Learning and Deep Learning are two new fields of study in the fields of information technology and agriculture. In India, agriculture is one of the most important occupations. As a result of a number of uncontrollable causes, our farmers face various challenges.

For good crop production, we must ensure that a specific crop can yield in a specific area and climatic condition. If a crop isn't producing as it should, it's most likely contaminated with a disease. So, our paper focuses on two parts: crop yield prediction, which will assist farmers in deciding which crop to plant, and crop leaf disease identification, which will assist farmers in quickly identifying the disease with a single click. We would be able to make more strategic crop production decisions with the aid of prediction. We can use machine learning to gain insights into the crop life cycle, which can be very useful.

Machine learning is an effective decision-making tool for forecasting crop yields, as well as determining which crops to plant and what to do during the growing season. Plant diseases are typically caused by rodents, insects, and pathogens, and if not addressed quickly, they can significantly reduce yield. A number of crop diseases are causing agriculturists to lose income. Crop diseases are a huge danger to food security, but due to a lack of competence in many regions of the world, quick detection is challenging. Thanks to a combination of expanding global technology penetration and recent breakthroughs in computer vision enabled by deep learning, smart technology assisted disease diagnosis is now conceivable. In the field of computer vision, detecting plant diseases is a critical research subject. It's a technique for taking pictures of plants with computer vision equipment in order to see whether they contain diseases or pests. Plant disease and pest detection equipment based on computer vision is being used in agriculture to replace conventional naked eye recognition.

The proposed framework has two stages: the first stage deals with training data sets, and the second stage deals with real-world data sets. This involves both stable and diseased data sets for training. The second step entails keeping an eye on the crop and determining the disease.

Keywords— Agriculture, Crop Production, Machine Learning, Decision-Making, Crop Diseases, Disease Detection, Deep Learning, Computer Vision, Training Data

I. INTRODUCTION

One of the most significant occupations practiced in India is agriculture. It is the largest economic area, with about 60 per cent of the country's land being used for agriculture to meet the needs of 1.2 billion people. Agriculture modernization is therefore very necessary and this will lead our country's farmers towards benefit. The earlier yield prediction was carried out by taking into account the experience of the farmer in a specific field and crop. Farmers, on the other hand, are pushed to raise more and more crops since conditions change swiftly from day to day. Many of them do not have appropriate knowledge of the new crops and are unaware of the benefits they gain when farming them, as is the case currently. In addition, through recognizing and predicting crop production in a variety of environmental circumstances, farm productivity can be improved.

It is possible to build an autonomous framework for disease classification of crops with the recent development in image processing and pattern recognition techniques. Firstly, track the plants annually. Photos of diseases are acquired by farmers using cameras or scanners. To interpret the image content by image processing processes, the acquired image must then be processed. A crucial point for crop management is the need for early pest/disease detection. Different procedures are carried out for this purpose, such as manual observation of plants. There are no precise measures provided by this method. Automatic detection is also very critical for disease detection at an early stage. Our aim is to build a system of detection that is versatile and easy to adapt to various applications. Traditional manual counting is tedious, time-consuming and subjective, since it relies on the skill of the observer. We propose to automate recognition and counting based on computer vision in order to solve these difficulties. Computer vision techniques are simpler to apply in our system to acquire high-resolution images of leaves.

I.A.1 PROBLEM STATEMENT

To learn and analyse the farming factors in order to acknowledge the farmers concern and come up with efficient ways to elevate the agricultural sector.

I.A.2 OBJECTIVE

To provide accurate crop prediction and crop disease detection based on data collected from the famers such as crop images, location along with the datasets that we have collected such as previous year crop yield, crops suitable for different soil type, weather parameters for previous years and crop leaf image dataset consisting of both healthy and diseased crop images.

II. SCOPE OF THE PROJECT

The aim of the project is to determine an area's crop yield by analysing a dataset that includes some significant or relevant features to crop production, such as temperature, moisture, rainfall, and previous crop production. To predict a continuous value, regression models are used. It is a supervised technique. The coefficients are pre-processed and fit into the trained data during training and construction the regression model. The main focus here is to reduce the cost function by finding the best fit-line. The output function facilitates in error measurement. During training period, error between the predicted and actual values is reduced in order to minimize error function. Python is used for this project. For a developing economy like India, agriculture is the cornerstone, and there is an immense need to preserve agricultural sustainability. It is, therefore, a valuable addition to the economic and agricultural welfare of countries around the world.
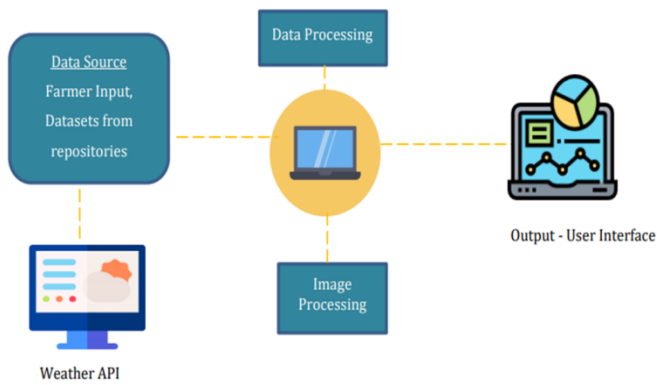
Fig .1 Agrovision Approach

III. PROPOSED SYSTEM

Agriculture is one of India's main game changers and a big money generator. Seasons, markets, and biological rhythms all influence crop production, yet fluctuations in these patterns cause farmers to lose money. This component can be reduced by employing an appropriate strategy based on understanding of soil types, pressure, acceptable weather, and crop kind. Weather and crop kinds, on the other hand, can be predicted using a helpful dataset that can assist farmers in anticipating the most profitable crops to cultivate.

- Analysing different factors on which crop yield depends season, land area, pests etc., using Machine Learning.
- Analysing the distribution of crops all over India based on past year data set and predicting the current production by using different algorithms.
- Provides guidance to farmers regarding new machines that can be used on fields.
- Keeps the farmers updated with the already existing government schemes or new or upcoming schemes.

The project will assist the farmers for selecting the crop properly at the earlier stage according to the climatic conditions of his farm. The proper selection of the crop at earlier stage will ultimately improve the crop yield and help to reduce the depression of the

farmer as well as there will be no need of re-cropping but it has its disadvantages –

- Existing system security is always at stake.
- Does not include accurate Crop Yield Prediction and Crop Disease Detection functionalities.
- Disease of crop can be detected with the help of the images of the crop, which does not exist in present system

IV. EXISTING SYSTEM

IV.A.1. A Climate-Smart Agriculture Weather Station
- A climate smart agricultural system is a means of monitoring the weather in a certain location and growing necessary crops in accordance with that climate.
- Uses real-time weather monitoring system
- This will assist farmers in growing the necessary quantity of crops on the required acreage and in determining the maximum and minimum precipitation temperatures in that area.

IV.A.2. Crop Yield Prediction Using Data Mining
- This paper proposes a system for prediction of production of crops in the current year. In order to determine the crop production, it uses a data mining algorithm K-Means.
- This system also employs a fuzzy logic-based prediction technique. Fuzzy logic is a rule-based prediction logic that is used to apply a set of rules to the land for farming, rainfall, and agricultural production. This study provides a comprehensive understanding of how K-Means may be used to examine data sets.
- We will apply the set of rules to anticipate which crop would return the most profit based on previous years' crop costs and current soil and weather data, similar to how they have applied the set of rules in the form of fuzzy logic.

IV.A.3. Image Processing Techniques for Early Pest Identification in Agriculture Crops

- The goal is to classify other stages of the white fly (eggs, larvae) and other bio aggressors (aphids) or plant diseases (powdery mildew). New artifacts for detecting or new image processing programs to retrieve the corresponding information may be implemented using a cognitive approach.
- Using Computer Vision and Algorithms of Artificial Intelligence.
- The collaboration of complementary disciplines and techniques that led to an automated, stable and versatile framework is demonstrated. For fast identification of white-flies, the prototype device proved effective.

Although a lot of research has been conducted for developing the decision support system for farmers, the most of the research focus on the crop management, crop disease management and crop yield forecasting. But the farmer's crop selection at the earlier stage is one of the most important factors since appropriate crop selection at the earlier stage will help farmers to improve crop management and crop yield. Crop forecast is a common issue that arises. A farmer was curious about how much output he should expect during the rising season. Previously, this yield estimate was based on a farmer's long-term expertise with specific yields, crops, and meteorological circumstances. Instead of worrying about crop forecast, farmers go straight for yield prediction with the current technique. Unless the correct crop is forecasted, the yield will be better, and pesticides, environmental and climatic parameters related to the crop will not be considered using existing methods.

V. DJANGO

Django is a Python-based open-source backend web application platform. The main goals are simplicity, versatility, reliability, and scalability. In Django, each function and component has its own name (for example, HTTP responses are called "views"). It also

has a more user-friendly admin panel than Laravel or Yei, as well as other technical features.

Fig .2 Django Features

VI. MACHINE LEARNING

Machine learning is a sort of artificial intelligence that allows software applications to improve their accuracy in predicting outcomes without being particularly programmed to do so. Various statistical and mathematical methods are used to measure the efficiency of ML models and algorithms. The trained model can be used after the conclusion of the learning process to identify, predict, or cluster new research data using the knowledge gained during the training process.

VII.   DEEP LEARNING FOR COMPUTER VISION

Computer vision is an area that involves making " see 'a machine. Instead of the human eye, this device uses a camera and computer to locate, monitor and quantify targets for further processing of images. Example of image processing includes:

- Normalizing the image's photometric attributes, such as brightness and colour.
- Cropping the bounds of the image, such as centring an object in a photograph.
- Removing digital noise from an image, such as digital artifacts from low light levels.

With the development of computer vision, such technology has been widely used in the field of

agricultural automation and plays a key role in its development. Deep learning in computer vision has made rapid progress over a short period. Some of the applications where deep learning is used in computer vision include face recognition systems, self-driving cars, etc. Convolutional neural networks, also known as convnets, a type of deep learning model universally used in computer vision applications. So, the objective of CNN is to perform two tasks: first is feature extraction and second is aggregating all the extracted features and making a prediction based on it.

VIII.   WEB DEVELOPMENT

A web application is a computer program that performs tasks over the Internet using web browsers and web technologies. A mixture of server-side scripts (PHP and ASP) are used by Web applications to manage information storage and retrieval, and client-side scripts (JavaScript and HTML) are used to display information to users. Online applications are normally encoded in languages supported by the browser, such as JavaScript and HTML, as these languages depend on the browser to make the program executable. Some of the apps are complex, requiring processing on the server side. Others are entirely static, with no server processing needed. A web application includes a web server to handle client requests, an application server to carry out the requested tasks, and often a database to store the information. Technology for application servers varies from ASP.NET, ASP and ColdFusion, to PHP and JSP.

IX. DATABASE

A database is a set of data that has been structured so that it can be easily accessed and maintained. To make it easier to locate relevant information, you can organise data into tables, rows, and columns, as well as index if. database handlers design a database such that all users have access to the data from a single collection of applications. The database's primary goal is to manage a vast volume of data by storing, extracting, and handling it. Databases are used to manage a large number of complex websites on the Internet today. Consider a model that tests the availability of hotel rooms. It's an example of a database-driven interactive website. Databases such as MySQL, Sybase, Oracle, MongoDB, Informix, PostgreSQL, SQL Server, and others are accessible.

X.   SYSTEM ARCHITECTURE

The System takes input from user i.e., user location, crop decided to grow and no. of hectares of land.

- Based on the location and chosen crop, system takes previous years data (soil attributes, rainfall, and weather) from the repositories and analyses the data and predicts which crop yields more or is more preferred to grow based on datasets of weather, soil and more.
- For crop disease detection, it takes that crop images as input from the user. Based on the image processing model it classifies the crop disease and also comes up with best possible solutions for it.
- Data Source – User Input, Weather API, Datasets from repositories
- Data Processing – ML Model consisting KNN, SVM, decision tree, Random Forest.
- Image Processing – Convolutional Neutral Networks.
- Results - User Interface
- It keeps the farmers up to date with the new farming equipment, technologies and strategies to increase the yield.
- From sowing seeds to harvest, Agrovision will assist farmers in all stages of farming

USE CASES IN SYSTEM
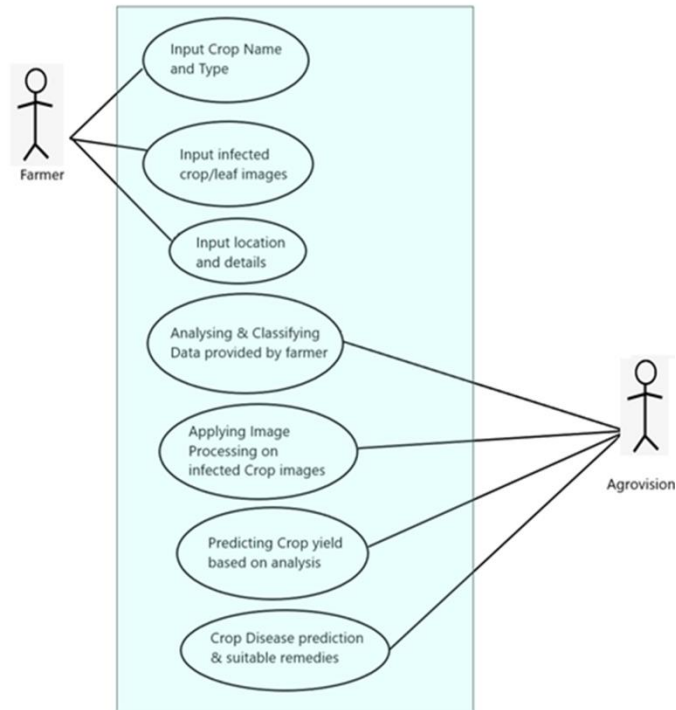
Fig .3 Use cases of Agrovision

DATA FLOW DIAGRAM
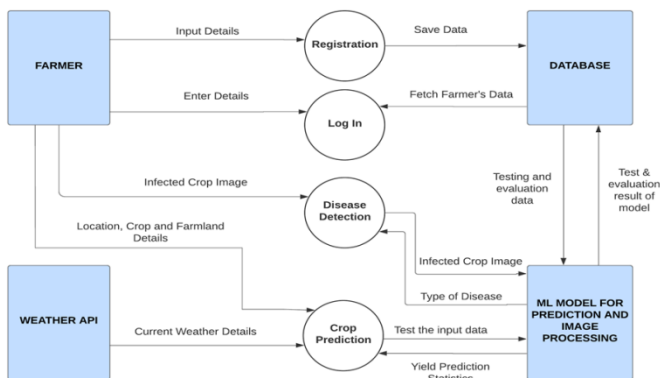
Fig .4 Data Flow Diagram of Agrovision

ACTIVITY DIGRAM
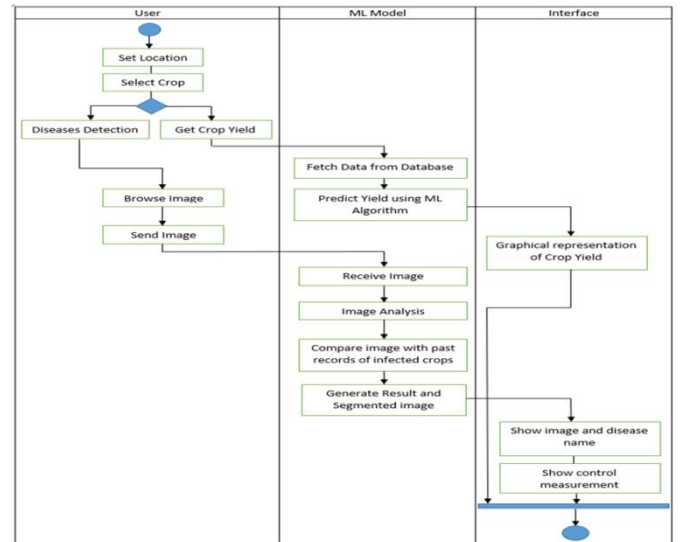
Fig .5 Activity Diagram of Agrovision

FLOW CHART
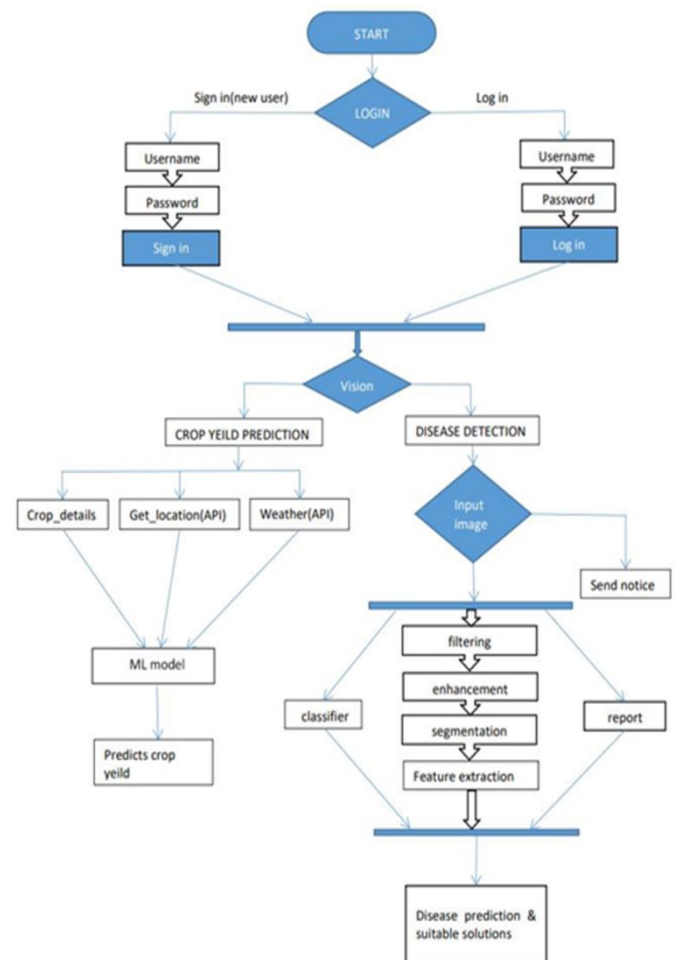
Fig .6 Flowchart of Agrovision

XI. FUTURE WORK

As the methodologies of smart farming increase, there will be a huge requirement for the implementation of newer technologies. The framework can be expanded to help farmers by uploading the farming picture to the mobile application. The efficiency of pre-processing is restricted by the amount of unwanted data, such as leaves and grass, etc. Because of this undesirable information present in the input picture, it can cause problems both during training and classification. The image parameters used for study, such as climate factor, moisture, and past data collection. In the future, IOT will be used to connect all farming devices over the Internet. The sensors can be used in farming to collect information on current farm conditions, and devices can increase moisture, acidity, etc. accordingly. The vehicles used in farm like tractor will be connected to internet in future which will, in real time pass data to farmer about crop harvesting and the disease crops may be suffering from thus helping the farmer in taking appropriate action. Further the best profitable crop can also be found in light of the monetary and inflation ratio.

XII. CONCLUSION

Agriculture has always been the most important sector for survival. There are a lot of difficulties faced by our farmers these days due to various unpredictable reasons. Hence, as engineers, we need to collaborate with farmers and provide them a solution to improve the quality and quantity of crops. Our project is the first step towards it. Prediction can help us make strategic decisions in crop production. With machine learning, we get insights about the crop life which can be very beneficial. Our model has learned with the train datasets and test datasets to produce optimal solution. We are given input as farmers current farming land location, farmer decided crop, and number of hectares of land in the application. Our system analysed the data and produced the predicted profitable crop and its required fertilizers list, overall yield per hectare and also shows total value of the crop based on current market price.

XIII. REFERENCES

[1]. https://conductscience.com/basic-tools-and-techniques-of-data-science/

[2]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm

[3]. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[4]. Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output. 2018 27th Wireless and Optical Communication Conference (WOCC).

[5]. Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C.,Basurto,C.(2015). "Crop-planning, making smarter agriculture with climate data."

[6]. Harvey, C. A., Rakotobe, Z. L., Rao, N. S., Dave, R., Razafimahatratra, H., Rabarijohn, R. H., et al. "Extreme vulnerability of smallholder farmers to agricultural risks and climate change.

[7]. "Plant disease: a threat to global food security" Strange, R. N., and Scott, P. R.

[8]. Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks,"

A P2P Botnet Detection Technique Using Machine Learning Classifiers

Yash Patwa[1], Tulika Kotian[1], Ralin Tuscano[1], Ms. Alvina Alphonso[1], Dr. Nazneen Ansari[2]

[1]Department of Information Technology, University of Mumbai, Mumbai, Maharashtra, India

[2]Department of Computer Engineering, University of Mumbai, Mumbai, Maharashtra, India

ABSTRACT

Today, botnets prove to be one among many scandalous perils to security in networks. While Client-Server botnets employ a centralized communication architecture, Peer-to-Peer(P2P) botnets acquire a decentralized structure for trafficking commands and controlling data, hence making them more difficult to be identified in a network. In this paper, the authors propose an effective system to detect Peer-to-Peer botnets by applying machine learning algorithms to network traffic parameters. The data from the CTU-13 dataset is input into the system. The proposed system has 3 phases. In the first stage, feature reduction was performed on the network traffic to recognize which of the features affected the classification considerably. In the second stage, the detection model was developed, which classified the traffic into Botnet(malign) traffic and Legitimate(benign) traffic in the last phase. The output of the system generates the classification of the network traffic with visualizations to gain insights into the network activity. The five machine learning algorithms employed are Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Logistic Regression, and Naive Bayes. On performing comparative analysis, the Decision Tree algorithm successfully detected Peer-to-Peer botnet traffic by demonstrating an accuracy of 99.90%.

Keywords — botnet detection, decision tree algorithm, machine learning, network security, P2P botnets, ReactJS

I. INTRODUCTION

In today's world of a billion Internet-connected devices, security issues are increasing as the network environments have become more intricate. Conficker was one of the largest botnets so far that affected 10.5 million computers. The damage caused on such a scale is huge and for the hackers, more damage means more profit. This had a huge impact on various sectors including government organizations, large institutions, and almost every social networking website like Facebook, Twitter, Instagram, etc., e-commerce websites Amazon, Flipkart, etc., in short, every firm on the internet was compromised by this malware.

A Botnet is a network of robots used for committing a cybercrime on the internet. A botnet mainly consists of three elements that are mandatory for performing malicious activity successfully: attackers, bots, and handlers [5]. The cybercriminals controlling the botnets are called Botmasters. Centralized botnets can be detected and destroyed. However, P2P botnets are difficult to detect and destroy because of their cagey nature [4]. Thus, P2P botnets are becoming popular among the attackers.

Today more and more businesses are coming online, thus securing businesses against physical and data threats is becoming very important. Now, with many devices communicating with each other over wired, wireless, or cellular networks, more complex and distributed networks are into the picture. Today, mobile botnets have become very popular due to the increased use of smartphones [2]. Mobile botnets pose serious threats to mobile security. Hence, network security becomes a very important concept. Classification in machine learning is a supervised machine learning approach in which the model is trained with labeled data which helps the model to classify new observations based on what it has learned from the data.
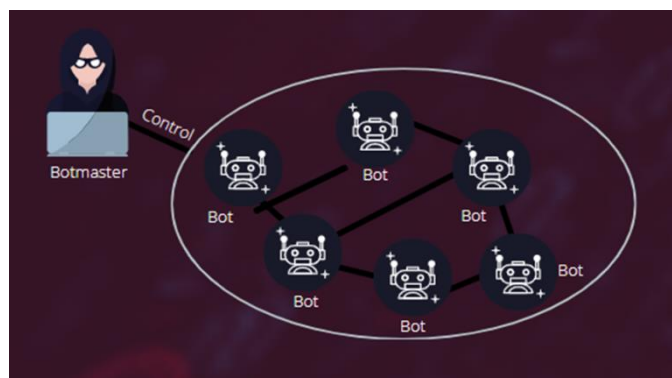
Fig. 1 P2P Botnet Structure

Due to the increase of IoT devices in networks, network security is getting more convoluted. Botnets will continue to be one of the major reasons for jeopardy in cybersecurity. Moreover, these kinds of malicious software are available in the market for free leases. Hence, a model to protect systems from one of the popular malware - botnets, has been proposed. This system proposes the development of a multi-phase detection of botnets that will make use of machine learning classifiers like Decision tree, SVM, Logistic Regression, etc.

II. RELATED WORK

Paper [1] proposed the implementation of a host-based intrusion detection system for botnet detection.

The authors have used a variation of a genetic algorithm to detect the peculiarity of these attacks. The experimental results concluded that a fitness level below 65% and a fitness level above 85% should be avoided as it leads to false-positive results.

Paper [2] provides a review of botnet trends, their evolution, and several ways to mitigate them. The authors have discussed the threats posed by the botnets viz. DDoS attack, Data theft, Spam, and more. The authors provide preventive measures viz. update the device to the latest operating system, avoid clicking on suspicious links, etc. The authors suggest detective measures like using the signature-based detection method, monitoring the anomalies in the network through DNS monitoring, Wireshark, and use of commercial network traffic monitors. The authors have also discussed the methods based on neural networks, logistic regression models, and machine learning with their respective accuracy of each method.

Paper [3] proposed "The Gunner System" which is a filtering approach and involves detecting DNS-based botnets. Through this approach, the authors aim to enhance the accuracy of DNS-based botnet detection. The authors conclude that the approach could successfully identify abnormal DNS queries and abnormal DNS responses.

Paper [4] uses feedforward artificial neural networks on convolutional features to detect P2P botnets using the idea of CNN. They achieved a detection accuracy of 94.7% and a false-positive rate of 2.2%. They have provided additional confidence testing for increasing the accuracy using the decision tree algorithm. After applying the confidence testing, the detection accuracy increases to 98.6% with a decreased false positive rate.

Paper [5] presents a thorough review of botnets, their lifecycle, and their types. Various botnet detection techniques like Signal Processing Technique, Entelecheia, PeerFox, Malicious Fast Flux Network Identification, Resource Sharing, and Online-Offline Detection were analyzed. The authors finally

conclude that these techniques provide high detection accuracy with negligible false positives. However, the authors also state that these techniques have a limited scope and cannot address all the problems of P2P botnets.

Paper [6] focuses on the research achievements of botnet detection that have been possible due to machine learning technology. The authors have explained the application process of machine learning in botnet detection. The security aspects of the existing solutions and the commonly used machine learning algorithms like SVM, KNN, etc. are analyzed and summarized.

Paper [7] focuses on the detection of social botnets in an online social network like Twitter. The proposed system uses a semi-supervised technique of machine learning for classification. The results of the experiments showed that they could accurately detect social bots with a low false-positive rate and an admissible detection time. One class SVM algorithm was used for the proposed system. The accuracy obtained was 99.95% with no false alarm rates.

Paper [8] makes use of the algorithms of artificial immune systems for detecting botnets in campus area networks. The proposed approach presents improvements in the BotGRABBER system. The experimental results indicated an accuracy rate of 95% with a false-positive rate of 3-5%. The model could successfully detect the HTTP, IRC, DNS, and P2P botnets using the clonal selection algorithm.

Paper [9] proposes the use of Bidirectional Long Short Term Memory based Recurrent Neural Network (BLSTM-RNN) along with Word Embedding for botnet detection. This model was also compared to a unidirectional LSTM-RNN. The experimental results showed that both the models returned high accuracy for the four attack vectors. The four attack metrics include the - Mirai, UDP, ACK, and DNS with corresponding accuracy rates 99%, 98%, 93%, and 98% respectively.

Paper [10] proposed a complete overview of existing botnet methods and along with a comparative analysis

on them. The authors discuss botnet detection techniques like honeynets, signature-based detection, anomaly-based detection. They have provided a comparative study of various botnet detection approaches with their respective design details, detection strategy, and limitations.

Paper [11] proposed an algorithm for botnet detection based on the idea of statistics using random walks and then validated it on real-world data of unstructured P2P botnets. The authors imitated malware spread on large network graphs with actual botnet data. Through the experimental analysis, the authors could conclude that their algorithm yielded a high-precision rate of 90%.

Paper [12] proposes a novel method to enhance IDS performance in botnet detection. Their technique makes use of dual statistical ways - low variance filter and Pearson correlation filter, during the process of feature selection. The feature reduction stage reduced attributes to be processed by the IDS system from 77 to 15, hence lowering the computational time and giving an accuracy of 97%. The experimental results show that, despite less number of features, the accuracy does not vary largely.

Paper [13] discusses shortcomings of merging smaller datasets to form large datasets, about how they degrade the prediction performance of the machine learning models They have made use of the PCA, TrAdaBoost algorithms for the detection. The authors suggest using transfer learning rather than traditional machine learning ways to intensify the performance of systems used for Botnet identification and detection.

Paper [14] uses behavior-based techniques for TCP/HTTP botnet detection because of their variant nature. The proposed system first reduces the traffic to remove the irrelevant traffic flows and reduce the traffic workload. After this, feature extraction is performed that uses comparative features like one-way connection density, the ratio of TCP packets, etc to efficiently identify DDoS attacks. The author has used PSO and SVM algorithms for classification.

Paper [15] presents a comprehensive examination of all network connection features in botnet creation and working. The authors have inspected protocols, botnet topology, and examined a set of highly sophisticated botnets existing today. Based on the analysis, they introduce a novel classification of generalized patterns for botnet communication using modeling diagrams.

Paper [16] is inspired by IP tracking technology. The authors have proposed an inventive botnet detection method that studies data packets using graph clustering. This technique analyzes the content of the packets with the timestamps of the traffic flow. This is facilitated by refining the HEMST clustering algorithm. On analysis, results showed that the correct rate in clustering could reach a high of 97%.

Paper [17] provides a summary of various machine learning techniques and their role in botnet detection. The main aim of the paper was to clearly define the contribution of machine learning algorithms in botnet detection. The authors have discussed anomaly and DNS-based detection techniques. However, they are of the view that ML-based techniques are the most effective ones.

Paper [18] introduces an original flow-based detection system that employs supervised machine learning to identify botnet traffic. On experimental analysis, their results indicated that the system could accurately detect botnet traffic using purely flow-based analysis on traffic with supervised techniques. Along with this, they also concluded that to achieve better results, the packet flow needs to be observed for a fixed time period and fixed packet rate.

Paper [19] proposes a scalable system that is capable of detecting sneaky P2P botnets. The system recognizes all nodes that are likely to be a part of P2P communications. It then gains statistical insights to identify P2P traffic and further differentiate between P2P botnet and non-botnet P2P traffic. Substantial gauging had shown high accuracy in detecting these botnets and impressive scalability.

Paper [20] reviews the ongoing research on HTTP-based botnet detection along with its pros and cons. They also propose an approach to better the HTTP-based botnet detection, with detecting HTTP bots with random traces. They successfully showed that their proposed method led to a favorable result of reducing false alarm rates in HTTP botnet detection.

III. PROPOSED SYSTEM

In the initial age of botnet activity, various approaches to expose botnets have been suggested based on the behavior botnets depict [10]. McDermott et al. stated that contemporary techniques of botnet detection such as flow-based or signature anomaly intrusion detections have been manifested unsuccessful in fending off the growth of botnets in IoT networks. This has been mainly because of the modifications in simple code hence providing outdated attack signatures or a lack of support from protocols (Sflow, NetFlow) in networks[9]. This paper is directed towards inspecting strategies and attributes to recognize distinct botnet behavior leading to better detection of botnets. The authors' aimed to use a powerful method to classify P2P network traffic and recognize botnet activity by performing analysis on distinct network features, followed by feature selection to take out irrelevant characteristics, and then a machine learning classification algorithm to classify network traffic into legitimate and botnet traffic.
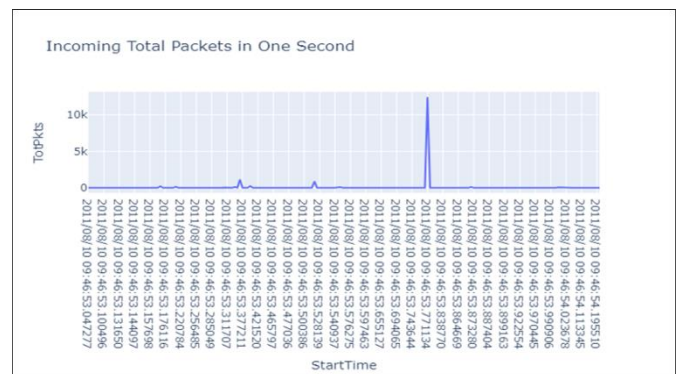
Fig. 2 Frequency and number of the influx of packets in a second

The evaluation of the system is limited to five algorithms for the comparison of experimental results namely Decision Tree, K-Nearest Neighbours, Support Vector Machines, Logistic Regression, and Naive Bayes. The accuracy of the finalized decision tree model is 99.9%. The presumption while applying machine learning approaches is that botnets create divisible patterns in network flow [18].

After comparison of the algorithms, the decision tree algorithm gave us optimum results, which is a supervised learning algorithm used in classification problems. Decision trees use a tree rendition approach to resolve problems where leaf nodes refer to a class label and features are portrayed on internal nodes. Decision trees are said to require less effort for data preparation during pre-processing.

A. System Architecture

The input to the model is a pre-labeled dataset i.e. CTU-13 dataset, that is publicly available and was captured in the CTU University. This dataset contains captures from different botnet scenarios. The model classifies the connection as legitimate or botnet. Firstly, the dataset is fed to the system as input. This input dataset contains various network attributes like the duration of the transmission, the number of bytes or packets transmitted, the protocol, the direction of the flow, and the label of the transmission. The training phase involves feeding this dataset i.e. Botnet Traffic, Normal Traffic, & Background Traffic. Next, the flows with fine-grained features are extracted and reduced to the effective features of the dataset. Feature selection was performed manually by checking how the features are related to one another by generating a correlation matrix with a heatmap between the features. Finally, the optimum classifier model (using a Decision Tree) is generated. The classifier model then classifies this traffic into two classes/labels of non-botnet and botnet connections. The output of the system generates the class of the network flow(malign/benign). The application also generates visualizations based on the input network

data flow which can give deep insights and provide an easier understanding to naive users on their network activity.
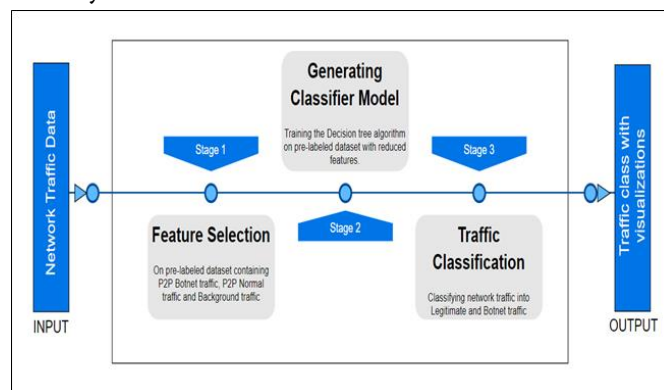
Fig. 4 System architecture

B. Algorithms Used

The classifiers first perform feature reduction and selection by selecting the most appropriate and useful features for the classification. At the first stage, feature reduction is performed on the network traffic to recognize which of the features affected the classification considerably. At the second stage, the detection model was developed, which classified the traffic into Botnet traffic and Legitimate traffic in the last phase. The proposed method gives an average accuracy of 93% considering all algorithms.

The algorithms used to provide a comparative analysis were Decision Tree, Support Vector Machine, K-Nearest Neighbour, Logistic Regression, and Naive Bayes algorithms. On the performance of these algorithms on this dataset and target prediction, Decision Tree gave exemplary results with a 99.90% accuracy among others. Also, decision tree algorithm models can handle sizable datasets, which is vital in this use case as substantial sizes of data packets flow in a network[12]. Naive Bayes gave the least accuracy of 72.25%.
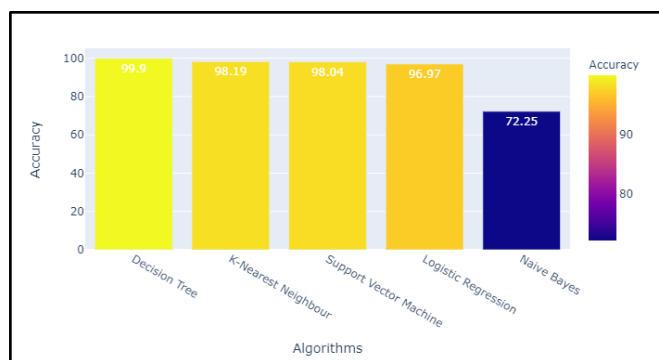
Fig. 5 Accuracies obtained by algorithms in detection model

IV. IMPLEMENTATION OF PROPOSED SYSTEM

The proposed model was trained and tested on the CTU-13 dataset. This dataset contains a large capture of mixed data containing botnet traffic, legitimate traffic, and background traffic from 13 scenarios like click fraud, fast flux, port scan, Distributed Denial of Service (DDoS), etc.

The CTU-13 dataset contains various features like the start time of the flow, duration of the flow, source and destination addresses, total packets, and bytes per transmission, with the corresponding labels of botnet activity (Normal, Background, or Botnet). The dataset was first analyzed using Wireshark and was then exported as an Excel file, to be analyzed using pandas, seaborn, matplotlib, etc.

During the analysis of features in the dataset, it was noticed that some features/attributes of network flow contribute distinctly to the nature of the traffic being botnet or non-botnet. A few of these features were the protocol of the network flow(TCP/UDP/ICMP), the ID that uniquely recognizes each transmission(since ID number ranges close to each other indicated similar nature of the network flows), etc.
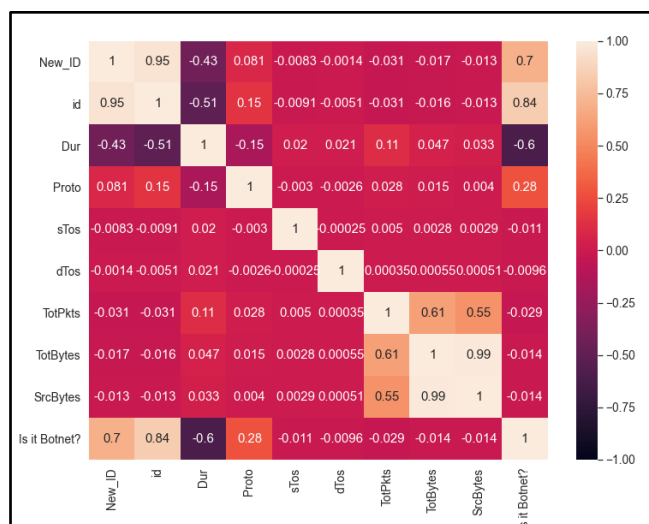
Fig. 3 Correlation of features of the dataset

The model was trained by splitting the data into training and testing sets, and five algorithms were tested on the system, where Decision Tree gave the most optimum results with minimum false rates.

The system was then deployed as a web application, using ReactJS technology. In this application, the user is first prompted to drop their network file (.binetflow) file which will contain the network flows and transactions of a fixed period. On processing this file, the system generates a dashboard for the user, wherein the risks to the user's network are mentioned by classifying the amount of botnet activity, and certain visualizations are generated via Tableau that produces statistics of the network flow.
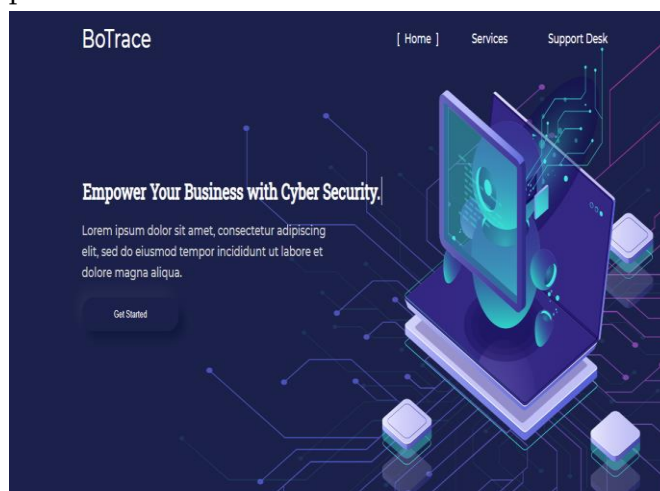
Fig. 6 Model deployed as a web application
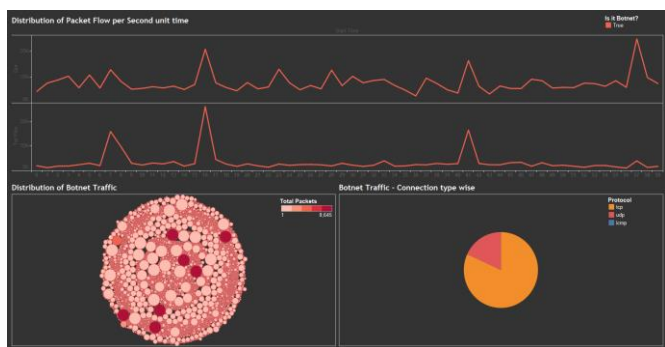
Fig. 7 Real-time statistics of network traffic generated from the system

V. RESULTS AND CONCLUSIONS

Bots use very minimal computing power to avoid disrupting or harming the device and thus alarming the user. Botnet designs continue to evolve which would make it even harder to detect.

The proposed system was successful in detecting the P2P botnet traffic. The multilayer approach overcomes the class imbalance problem of the single-layer botnet detection methods. The internet traffic is first reduced wherein only the TCP packets are filtered. Then the traffic is sent to the P2P and the Non-P2P traffic classifier where the traffic is filtered based on data packets, data stream, and session layer. The next step involves reducing the features that would marginally affect the classification. Finally, the traffic is fed to the machine learning model which then classifies it as legitimate P2P traffic and botnet P2P traffic.

The results of the proposed framework were compared with five different classifiers out of which Decision Tree gave an accuracy of 99.90% with minimum false alarm rates.

Hence, this system can be a useful tool for a network administrator or anyone who wishes to keep a check on their network flow, to track the presence of any malicious activity. In the future, the scope of this project can be further extended to detect other malware threats to a network like spyware, adware, worms, etc.

TABLE I

PERFORMANCE OF THE SYSTEM WITH RESPECT TO THE VARIOUS ALGORITHMS

Algorithm	Accuracy
Decision Tree	99.90%
K-Nearest Neighbour	98.19%
Support Vector Machine	98.04%
Logistic Regression	96.97%
Naive Bayes	72.25%

VI. REFERENCES

[1]. Y. ALEKSIEVA, H. VALCHANOV, and V. ALEKSIEVA, "An approach for host-based botnet detection system," 2019 16th Conference on Electrical Machines, Drives and Power Systems (ELMA), Varna, Bulgaria, 2019, pp. 1-4.

[2]. T. Lange and H. Kettani, "On Security Threats of Botnets to Cyber Systems," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2019, pp. 176-183.

[3]. K. Alieyan, M. Anbar, A. Almomani, R. Abdullah and M. Alauthman, "Botnets Detecting Attack Based on DNS Features," 2018 International Arab Conference on Information Technology (ACIT), Werdanye, Lebanon, 2018, pp. 1-4.

[4]. S. Chen, Y. Chen and W. Tzeng, "Effective Botnet Detection Through Neural Networks on Convolutional Features," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 372-378.

[5]. H. Dhayal and J. Kumar, "Botnet and P2P Botnet Detection Strategies: A Review," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 1077-1082.

[6]. X. Dong, J. Hu and Y. Cui, "Overview of Botnet Detection Based on Machine Learning," 2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Huhhot, 2018, pp. 476-479.

[7]. A. Dorri, M. Abadi, and M. Dadfarnia, "SocialBotHunter: Botnet Detection in Twitter-Like Social Networking Services Using Semi-Supervised Collective Classification," 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Athens, 2018, pp. 496-503.

[8]. S. Lysenko, K. Bobrovnikova and O. Savenko, "A botnet detection approach based on the clonal selection algorithm," 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kiev, 2018, pp. 424-428.

[9]. C. D. McDermott, F. Majdani and A. V. Petrovski, "Botnet Detection in the Internet of Things using Deep Learning Approaches," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1-8.

[10]. G. Khehra and S. Sofat, "Botnet Detection Techniques: A Review," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 1319-1326.

[11]. D. Muhs, S. Haas, T. Strufe and M. Fischer, "On the Robustness of Random Walk Algorithms for the Detection of Unstructured P2P Botnets," 2018 11th International Conference on IT Security Incident Management & IT Forensics (IMF), Hamburg, 2018, pp. 3-14.

[12]. F. A. Saputra, M. F. Masputra, I. Syarif, and K. Ramli, "Botnet Detection in Network System Through Hybrid Low Variance Filter, Correlation Filter and Supervised Mining Process," 2018 Thirteenth International Conference on Digital Information Management (ICDIM), Berlin, Germany, 2018, pp. 112-117.

[13]. B. Alothman and P. Rattadilok, "Towards using transfer learning for Botnet Detection," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, 2017, pp. 281-282.

[14]. A. Kapre and B. Padmavathi, "Behavior-based botnet detection with traffic analysis and flow intervals using PSO and SVM," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 718-722.

[15]. G. Vormayr, T. Zseby and J. Fabini, "Botnet Communication Patterns," in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2768-2796, Fourth quarter 2017.

[16]. X. Kong, Y. Chen, H. Tian, T. Wang, Y. Cai, and X. Chen, "A Novel Botnet Detection Method Based on Preprocessing Data Packet by Graph Structure Clustering," 2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Chengdu, 2016, pp. 42-45.

[17]. S. Miller and C. Busby-Earle, "The role of machine learning in botnet detection," 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, 2016, pp. 359-364.

[18]. M. Stevanovic and J. M. Pedersen, "An efficient flow-based botnet detection using supervised machine learning," 2014 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, 2014, pp. 797-801.

[19]. J. Zhang, R. Perdisci, W. Lee, X. Luo, and U. Sarfraz, "Building a Scalable System for Stealthy P2P-Botnet Detection," in IEEE Transactions on Information Forensics and Security, vol. 9, no. 1, pp. 27-38, Jan. 2014.

[20]. M. Eslahi, H. Hashim and N. M. Tahir, "An efficient false alarm reduction approach in HTTP-based botnet detection," 2013 IEEE Symposium on Computers & Informatics (ISCI), Langkawi, 2013, pp. 201-205.

Face Mask Detector with Deep Learning and MobileNetV2

C B Sri Sai Maheswari[1], Hema Surya[1], Lavanya G[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The corona virus COVID-19 pandemic is causing a global health crisis so the effective protection method is wearing a face mask in public areas according to the World Health Organization (WHO). The COVID-19 pandemic forced governments across the world to impose lockdowns to prevent virus transmissions. Reports indicate that wearing facemasks while at work clearly reduces the risk of transmission. An efficient and economic approach of using AI to create a safe environment in a manufacturing setup. A hybrid model using deep and classical machine learning for face mask detection will be presented. A face mask detection dataset consists of with mask and without mask images, we are going to use OpenCV to do real-time face detection from a live stream via our webcam. We will use the dataset to build a COVID-19 face mask detector with computer vision using Python, OpenCV, and Tensor Flow and Keras. Our goal is to identify whether the person on image/video stream is wearing a face mask or not with the help of computer vision and deep learning.

Keywords –- OpenCV, Tenser Flow, Keras, Computer Vision

I. INTRODUCTION

The trend of wearing face masks in public is rising due to the COVID- 19 corona virus epidemic all over the world. Before Covid-19, People used to wear masks to protect their health from air pollution. While other people are self-conscious about their looks, they hide their emotions from the public by hiding their faces. Scientists proofed that wearing face masks works on impeding COVID-19 transmission. COVID19 (known as corona virus) is the latest epidemic virus that hit the human health in the last century. In 2020, the rapid spreading of COVID-19 has forced the World Health Organization to declare COVID- 19 as a global pandemic. More than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas[6]. The corona virus epidemic has given rise to an extraordinary degree of worldwide scientific cooperation. Artificial Intelligence (AI) based on Machine learning and Deep Learning can help to fight Covid-19 in many ways. Machine learning allows researchers and clinicians evaluate vast quantities of data to forecast the distribution of COVID-19, to serve as an early warning mechanism for potential pandemics, and to classify vulnerable populations[1]. The provision of healthcare needs funding for emerging technology such as artificial intelligence, IoT, big data and machine learning to tackle and predict new diseases. In order to better understand infection rates and to trace and quickly detect infections, the AI's power is being exploited to address the Covid-19 pandemic. People are forced by

laws to wear face masks in public in many countries. These rules and laws were developed as an action to the exponential growth in cases and deaths in many areas. However, the process of monitoring large groups of people is becoming more difficult[5]. The monitoring process involves the detection of anyone who is not wearing a face mask. Here we introduce a mask face detection model that is based on computer vision and deep learning. The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are wearing masks not wearing face masks. The model is integration between deep learning and classical machine learning techniques with opencv, tensor flow and keras. We have used deep transfer leering for feature extractions and combined it with three classical machine learning algorithms[1]. We introduced a comparison between them to find the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection.

II. MACHINE LEARNING

Machine learning (ML)is the study of computer algorithms that improve automatically through experience. Itis seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks[3]. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through

unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics[2].

Machine learning approaches are traditionally divided into three broad categories, depending on the nature of the "signal" or "feedback" available to the learning system:

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

Other approaches have been developed which don't fit neatly into this three-fold categorization, and sometimes more than one is used by the same machine learning system [6].

III. DEEP LEARNING

Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features. Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features[4].

The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning. Deep learning excels on problem domains where the inputs (and even output) are analog. Meaning, they are not a few quantities in a tabular format but instead are images of pixel data, documents of text data or files of audio data[7]. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction [5].

IV. PROPOSED SYSTEM

The proposed system focuses on how to identify the person on image/video stream wearing face mask with the help of computer vision and deep learning algorithm by using the OpenCV, Tensor flow, Keras and PyTorch library.

Approach-
1. Train Deep learning model (MobileNetV2)
2. Apply mask detector over images / live video stream

Data at Source
The majority of the images were augmented by OpenCV. The set of images were already labeled "mask" and "no mask". The images that were present were of different sizes and resolutions, probably extracted from different sources or from machines (cameras) of different resolutions[7].

Data preprocessing
Preprocessing steps as mentioned below was applied to all the raw input images to convert them into clean versions, which could be fed to a neural network machine learning model.
1. Resizing the input image (256 x 256)

2. Applying the color filtering (RGB) over the channels (Our model MobileNetV2 supports 2D 3 channel image)
3. Scaling / Normalizing images using the standard mean of PyTorch build in weights
4. Center cropping the image with the pixel value of 224x224x3
5. Finally Converting them into tensors (Similar to NumPy array)

MobileNetV2
MobileNetV2 builds upon the ideas from MobileNetV1, using depth wise separable convolution as efficient building blocks. However, V2 introduces two new features to the architecture:
1) Linear bottlenecks between the layers, and
2) Shortcut connections between the bottlenecks.

The typical MobilenetV2 architecture has as many layers listed below. In Pytorch we can use the model library in TorchVision to create the MobileNetV2 model instead of defining/building our own model. The weights of each layer in the model are predefined based on the ImageNet dataset[6]. The weights indicate the padding, strides, kernel size, input channels and output channels. MobileNetV2was chosen as an algorithm to build a model that could be deployed on a mobile device. A customized fully connected layer which contains four sequential layers on top of the MobileNetV2 model was developed[4]. The layers are
1. Average Pooling layer with 7×7 weights
2. Linear layer with ReLu activation function
3. Dropout Layer
4. Linear layer with Softmax activation function with the result of 2 values.

The final layer softmax function gives the result of two probabilities each one represents the classification of "mask" or "not mask".

Figure 1: Detection of "Mask" and "No Mask"

V. IMPLEMENTATION

We are still developing the face mask detector project. As of now we have implemented the code for detecting face, pre processing it, loading mobilenetv2 and converting images to numeric arrays. In coming days we'll be training the complete model which will be able to detect "mask" and "no-mask" faces as well.

1. Implementation of image loading, pre processing and converting images into numeric array

```
for img in os.listdir(path):
    img_path = os.path.join(path, img)
    image=load_img(img_path,
    target_size=(224, 224))
        image = img_to_array(image)
        image = preprocess_input(image)
        data.append(image)
        labels.append(category)
```

2. Loading MobilrNetV2 network
load the MobileNetV2 network, ensuring the head FC layer sets are
```
baseModel     =     MobileNetV2(weights="imagenet",
include_top=False,
input_tensor=Input(shape=(224, 224, 3)))
```

Face Mask Detection in webcam stream
The flow to identify the person in the webcam wearing the face mask or not. The process is two-fold.
1. To identify the faces in the webcam
2. Classify the faces based on the mask

Identify the Face in the Webcam
To identify the faces a pre-trained model provided by the OpenCV framework was used. The model was trained using web images.
Floating-point 16 version of the original Caffe implementation[5].
2.8 bit quantized version using Tensor flow The Caffe model in this face mask detector.There has been a lot of discussion around deep learning based approaches for person detection. This encouraged us to come up with our own algorithm to solve this problem. Our work on facemask detection comprises of data collection to tackle the variance in kinds of face masks worn by the workers. The face mask detection model is a combination of face detection model to identify the existing faces from camera feeds and then running those faces through a mask detection model[3].

VI. LITERATURE REVIEW

Covid-19 Facemask detection with Deep Learning and Computer Vision by Vinitha and Velantina from IRJET- Volume 07 and Issue 08 ; have proposed a hybrid model using deep and classical machine learning for face mask detection. They have identified whether the person on image/video stream is wearing a face mask or not with the help of computer vision and deep learning.

The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are wearing masks not wearing face masks. The model is integration between deep learning and classical machine learning techniques with opencv, tensor flow and keras. They have used deep transfer leering for feature extractions and combined it with three classical machine learning algorithms. They've also introduced a comparison between them to find the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection. The face mask detection model is a combination of face detection model to identify the existing faces from camera feeds and then running those faces through a mask detection model.

VII.   CONCLUSION

As the technology are blooming with emerging trends the availability so we have novel face mask detector which can possibly contribute to public healthcare. The architecture consists of MobileNet as the backbone it can be used for high and low computation scenarios. In order to extract more robust features, we utilize transfer learning to adoptweights from a similar task face detection, which is trained on a very large dataset. We used OpenCV, tensor flow, keras , Pytorch and CNN to detect whether people were wearing face masks or not. The models were tested with images and real-time video streams. The accuracy of the model is achieved and, the optimization ofthe model is a continuous process andwe are building a highly accurate solution by tuning the hyper parameters. This specific model could be used as a use case for edge analytics. Furthermore, the proposed method achieves state-of-the-art results on a public face mask dataset. By the development of face mask detection we can detect if the person is wearing a face mask and allow their entry would be of great help to the society.

VIII. REFERENCES

[1]. P. A. Rota, M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp,K. Maher, M.-h. Chenet al., "Characterization of a novel coronavirus associated with severe acute respiratorysyndrome,"science, vol. 300, no. 5624, pp. 1394–1399, 2003.

[2]. Z. A. Memish, A. I. Zumla, R. F. Al-Hakeem, A. A. AlRabeeah, and G. M. Stephens, "Family cluster of middleeast respiratory syndrome coronavirus infections,"New England Journal of Medicine, vol. 368, no. 26, pp.2487–2494, 2013.

[3]. Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is higher comparedto sars coronavirus,"Journal of travel medicine, 2020.

[4]. Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of governmentinterventions: A datadriven analysis,"Journal of medical virology, vol. 92, no. 6, pp. 645–659, 2020.

[5]. N. H. Leung, D. K. Chu, E. Y. Shiu, K.-H. Chan, J. J. McDevitt, B. J. Hau, H.-L. Yen, Y. Li, D. KM, J. Ipet al.,"Respiratory virus shedding in exhaled breath and efficacy of face masks."

[6]. S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the covid19pandemic,"The Lancet Respiratory Medicine, 2020.

[7]. Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Peiet al., "Masked facerecognition dataset and application,"arXiv preprint arXiv:2003.09093, 2020.[10]Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review,"IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212–3232, 2019.

Prediction of COVID-19 Cases using ARIMA

Samhitha Mudiam[1], Prajwal SV[1], Sushmitha GS[1], Dr. R. Jaya[2]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

[2]Assistant Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

COVID-19, a new type of coronavirus which gets its name due to the shape of the virus, i.e. the crown like structure, is creating havoc all around the globe. In India, initially it was infecting people of the 50+ age group, but in the recent times, people of 21-23 and 28-31 years of age are also getting affected due to COVID-19. It affects the persons respiratory system and a person succumbs to death due to breathing problem. Being able to accurately forecast when the outbreak will hit its peak would significantly diminish the impact of the disease, as it would consent administrations to amend their policy thus and plan forward for the preventive steps needed such as public health messaging, raising awareness of citizens and increasing the capacity of the health system.

This study investigated the accuracy of a variety of time series modeling approaches for coronavirus outbreak detection in ten different countries with the peak number of inveterate cases. The results demonstrate that, given data produced using actual testing for a small portion of the population, machine learning time series methods can learn and scale to accurately estimate the percentage of the total population that will become affected in the imminent.

Keywords— COVID-19, Pandemic, ARIMA, Machine Learning, Time Series, India, Forecast.

I. INTRODUCTION

The 2019 Covid contagion started in Wuhan, China in December 2019 also caused severe damage almost everywhere in the world. 2019nCoV or COVID19, commonly known as Coronavirus, is a highly contagious new virus belonging to the Coronavirus family, which is said to be transmitted from animals to humans. This virus can cause mild to severe respiratory illness and death. However, the premature cases show that the infection is severe compared to cases of other coronaviruses, such as SARS CoV and MERS CoV, which spread rapidly from person to person. The degree is low. This shows that 2019nCoV is extremely infectious. What others. There are deceptive differences in the epidemiological examination and the ability to detect cases of infection in different countries. Currently, the United States has reported the peak numeral of 2019nCoV infections, but cases in Spain, Italy, France, and Germany are increasing dramatically every day. China is the origin of the disease and the number of

cases currently received is very limited. The first case of coronavirus infection in India was reported in Kerala on January 30, 2020. The case was imported from Wuhan, China. In the initial stage, the spread was very slow and only 3 people tested positive in more than a month. However, this number started to increase exponentially after a month and continued.

In the current work, we use the ARIMA model to envisage the onset of 2019nCov disease. The ARIMA model is more suitable for predicting natural problems.

In our research, we identified the best ARIMA model and then predicted the number of cases in the next few days. The main purpose of this research is to seek out the simplest prediction model and apply it to the longer-term prediction frequency of COVID19 cases in India.

II. LITERATURE REVIEW

The paper on COVID-19 case prediction in India used time series data. For forecasting time series, ARIMA modeling is one of the best modeling techniques. The ARIMA model is always characterized by certain parameters, and the model is expressed as ARIMA (p, d, q). Here, p epitomizes the order of autoregression, d epitomizes the degree of trend difference, and q epitomizes the order of moving average. We have applied the ARIMA model to the time series data of inveterate COVID-19 cases in India. ACF and PACF are used to find the preliminary number of ARIMA models. These ARIMA models are then tested to determine the difference between normal and paperwork. Then verify its accuracy by looking at its MAPE, MAD and MSD values to determine the premium model to be predicted. In addition, use accurate measurement output to compare ARIMA best-fit models with linear trend, quadratic trend. MAPE, MAD, MSD, to select the best model for prediction. The thinnest model is the model with the lowest measured value of all. After fitting the model, estimate its parameters, and then verify the model.

In the following article, we propose a method to guesstimate the share of active cases in relation to the total population in ten countries using time series methods inspired by statistics and machine learning. To be more precise, six different time series methods were developed, namely ARIMA HWAAS, Facebook's Prophet` and then carried out for each of the following countries for comparison:

United States, United Kingdom, Italy, Spain, Russia and Brazil. In terms of evaluation indicators, the RMSE is used to evaluate the routine of each time series model. The results show that, although there is no single method to predict the cases of activity in different countries, ARIMA and TBAT have shown excellent performance in seven out of ten countries and have achieved excellent results in another two countries. The second-best result.

III. EXISTING SYSTEM

To identify the patterns and trends of various infectious disease-related events, machine learning and some statistical methods were previously used. Time series forecasting is part of the machine learning and statistical methods that have been used successfully for these infectious diseases. For example, modeling leptospirosis and its relationship to rainfall and temperature, and the temporal correlation between the monthly number of Plasmodium falciparum cases and the El Niño Southern Oscillation (ENSO).

IV. PROPOSED SYSTEM

Use ARIMA with a time series model for forecasting. There are several effective models for predicting time series, one of which includes the Autoregressive Integrated Moving Average (ARIMA) model; it was originally developed to calculate economic statistics. Its statistical properties, execution of the well- known Box-Jenkins method during model training, and its ability to perform various exponential smoothing

models have all contributed to its reputation and spread and achieved success worldwide. The working principle of

ARIMA is as follows. First, you will assume that there is a linear relationship between the two entities. One of them is the time series value, and the other is done by trying to use linear correlation in the observation, the purpose is to citation the homegrown pattern while confiscating the high frequency noise from the data. This method of brings us two main benefits. First, it provides us with high interpretability, because based on the assumptions of the model, it can predict the rapport between the dependent variable and the independent variable. The relationship is easy to understand and explain. This allows researchers not only to obtain information about the relationship between the current state and the past state (endogenous variables), but also to know the influencing inputs (exogenous variables) outside the series of states. The second benefit involves model selection. For ARIMA models, this can be done automatically to maximize the accuracy of the forecast. Another benefit of ARIMA models is that they can adapt to systems controlled by the dynamics of these systems over time. These changes can be made by bring up-to-date the model based on topical events to predict the imminent state of the system.

V. METHODOLOGY

Dataset-- Data is collected over all countries. Confirmed count of recovered and death cases only due to COVID-19 infection is placid from India as well as highly infected countries (US, Spain, Italy, France, Germany, China and Iran) and countries in South-East Asia region as per World Health Organization region classification, from the official website of Johns Hopkins University. This data is employed to create predictive models.

Model Development-- The built model is employed to conjecture confirmed COVID-19 cases for the next 20 days. The model for forecasting future confirmed COVID-19 cases is represented as,

$ARIMA\ (p,\ d,\ f): Xt = \alpha 1 Xt{-}1 + \alpha 2 Xt{-}2 + \beta 1\ Zt{-}1 + \beta 2 Zt{-}2 + Zt$

The trend of impending incidences is often assessed from the previous cases and a statistic analysis is performed for this persistence. Time series forecasting refers to the utilization of a model to forecast future data supported previously observed data. In the present study, time series investigation is used to recognize the trends in confirmed COVID-19 cases in India over the dated span of 22 January 2020 to 13 April 2020 and to predict future

cases from 14 April 2020 till 3 May 2020. The level of statistical significance is set at 0.05. A graph is plotted for actual confirmed cases and predicted confirmed cases with regard to time to verify the efficiency of the model. To get an idea of the recovery and death trends in India, a graph is plotted with respect to time. A comparative study is additionally performed to look at the status of confirmed COVID-19 cases of India with reference to those of highly infected countries. A similar comparison is formed with the countries of South-East Asia region also. All the model developments, computations and comparisons have been performed using Minitab software

VI. EXPECTED RESULTS

To support the prevention of this pandemic and to briefly understand the increase in daily cases and the number of recoveries and deaths, we developed a model to predict future COVID-19 cases. We pass the data of existing cases and the relevant time of recovery and death. The ARIMA model will obtain this data and apply an algorithm, which successively will predict the longer-term growth of the case, and can confirm the confirmed diagnosis constructed on the model. prediction the numeral of cases is about to surge and time series analysis also shows an exponential increase in cases of infection. We hope this predictive model helps governments and medical

personnel by issuing warnings about upcoming global threats. And it helps them take the necessary steps to stop the spread of COVID-19. Below are some snapshots of the results we got.
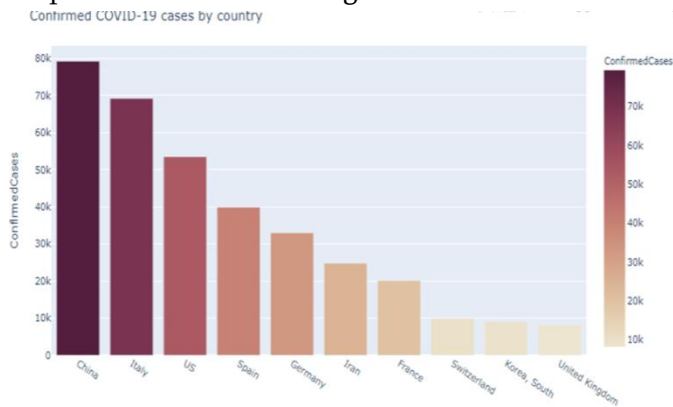
Fig. 1 Established COVID-19 cases by country

Fig. 2 Established COVID-19 cases per day in china.
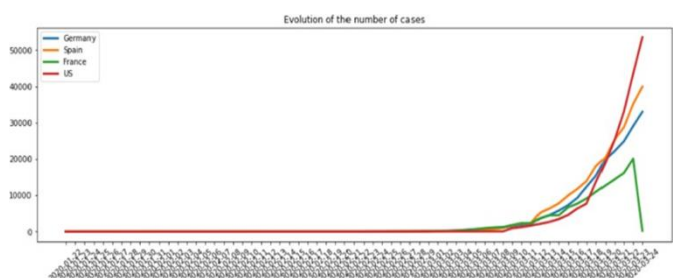
Fig. 3 Trend in the number of cases from January to march (2020)
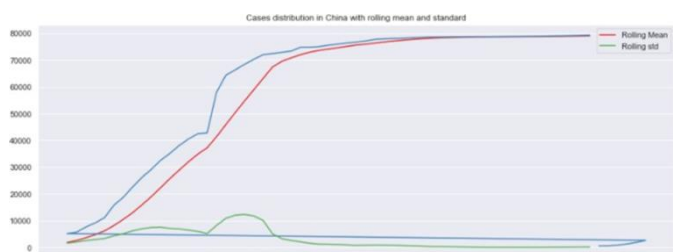
Fig. 4 Rolling mean and standard implications

```
                       ARIMA Model Results
==============================================================================
Dep. Variable:                    D2.y   No. Observations:                   23
Model:                   ARIMA(0, 2, 0)   Log Likelihood                -163.340
Method:                            css   S.D. of innovations            293.767
Date:                 Mon, 10 May 2021   AIC                            330.679
Time:                         16:45:26   BIC                            332.950
Sample:                              2   HQIC                           331.251

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        173.7609     61.255      2.837      0.005      53.704     293.818
```

Fig. 5 ARIMA model outcomes
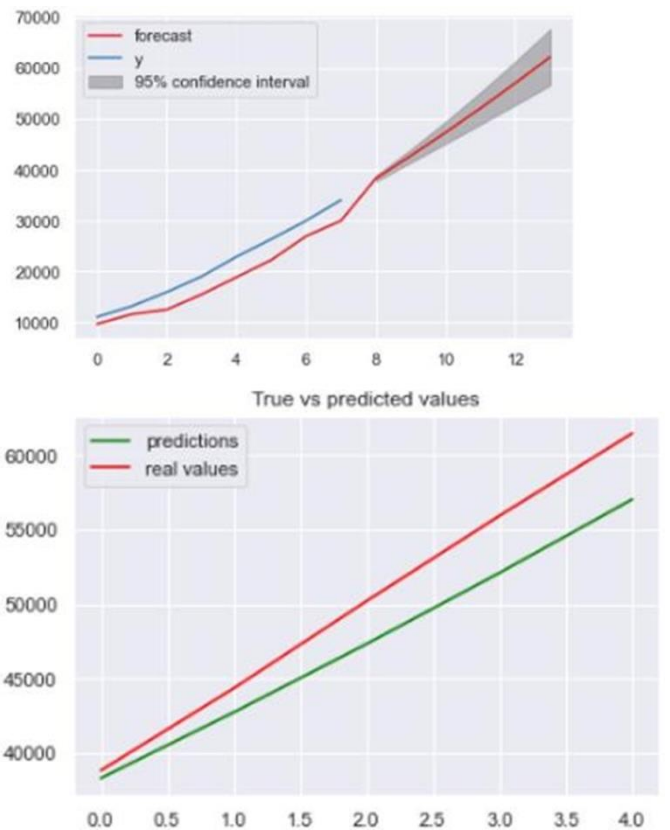
Fig. 6 Forecasted and Predicted Graphs

VII.   FUTURE SCOPE

Future amendments to supplementary advance the model's forecasting accurateness include creating a collection of proposed models that will combine the world's best models to reduce overall errors. This data can be related to general demographic or national characteristics that change very slowly over time, such as density or geographic attributes, but it can also be time series data, such as changes in temperature or humidity during a pandemic, or quantitative data. The time, severity, and duration of social distance measurement, such as air quality data

in different countries or regions over time. Another future goal is to use some form of learning transfer to bring learning outcomes from one country to another.

VIII. REFERENCES

[1]. COVID-19: A Comparisonof TimeSeries Methods toForecast Percentage ofActiveCasesper Populationhttps://www.mdpi.com/20763417/10/11/3880?type=c heck_update&version=1

[2]. World Health Organization, Coronavirusdisease (COVID-19) outbreak

[3]. Paules, C. I.; Marston, H. D.; Fauci, A. S. Coronavirus infections—more than just the common cold, JAMA 2020, 323, 707.

[4]. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 2020, 382, 727.

[5]. Fernandes, NEconomicEffectsoCoronavirus OutbreakCOVID-19) ontheWorlEconomy.2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3557504

[6]. Zhang, Y.; Yang, H.; Cui, H.; Chen, Q. Comparison of the Ability of ARIMA, WNN and SVM Models forDrought Forecasting in theSanjiangPlain,China.Nat.Resour. Res. 2019, 29, 144

[7]. IHMECOVID-19HealtServiceUtilitionForecasting Team. Forecasting COVID-19 impact onhospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv 2020. doi:10.1101/2020.03.27.20043752.

A Novel and Innovative Scheme for IoT Based Weather Monitoring System

Meghana.B[1], Savitha H.M[1]*, Bebi[1], Dr. Anidha Arulanandham[1]

[1]Department of Computer Science, New Horizon College of Engineering Outer Ring Road, Panattur post, Kadubeesanahalli, Bengaluru-560103, Karnataka, India

ABSTRACT

These days weather forecasting is unpredictable due to drastic changes in the weather. Because of that, Weather report is used more often for monitoring the changes in the climate and environmental conditions continuously. Internet of Things platform uses ThingSpeak which is open source IoT application and API which will continuously stores and displays the weather parameters such as temperature, humidity, wind speed, light intensity etc., and these information will be monitored continuously using an user interface with Arduino IDE. This can be done with the help of a microcontroller communications using Wi-Fi hotspots. Satellites fail to give exact conditions of the weather conditions of some specific places. At that time, using Weather Monitoring System with all the parametric sensors can be controlled by ESP32 microcontroller as the server which collects all the data sent by sensors, which will be sent to the database by ThingSpeak. As a result the weather data of any particular area around the world, can be observed. The data collected can also be saved in google sheet format by IFIT tool to analyze the data easily. This system monitors the changes in the weather conditions and provides users fastest way to access the information from anywhere.

Keywords — Internet of Things (IoT), Arduino IDE, Wi-Fi hotspot, Microcontroller, Weather Monitoring

I. INTRODUCTION

Climate plays a vital role in human life. The unprecedented growths of industries and vehicular traffic have seriously affected the purity of clean air and environment [1]. Satellite Weather report system gives status of present climatic and environmental conditions which doesn't provide the exact condition of all places across the world. The building sector offers a great potential for the energy savings, where it is necessary to have accurate weather data of the exact location where the building is being built in-order to improve the calibration of energy simulation programs [2]. By developing a control local weather reporting system with ESP32 and wemos D1 mini microcontroller can minimize the error in weather forecast system at exact locations. A precision agriculture and farming can be defined as the art and science of using technology to improve crop production [3]. Even though there is water scarcity overall, 50% of water is wasted in agriculture due to improper scheduling of irrigation [4].

At present, most of the technologies mainly focus on controlling and monitoring of different activities. These are increasingly emerging to reach human needs [5]. Most of these technologies are focused on efficient monitoring and controlling system is required to monitor and assess the conditions in case

of exceeding the prescribed level of parameters [5].The changes in the environment affects animals, plants and human beings and these can be monitored and controlled by smart environmental targets [5]. By using the embedded intelligence into the environment makes the environment interactive with other goals, this is one of the applications that smart environment targets [5].

There is a necessity in security or alarming system that gives warning and alerting when there is a bad condition at the place. The existing technologies are developed using micro controllers like Arduino, Node MCU etc. and ARM processors like Raspberry Pi [6]. So, machine learning techniques achieved better performance than traditional statistical methods in learning without being expressly customized [7]. Data measured by the stations could be used for various purposes, such as air quality management to reduce pollutant gases in the local atmosphere and climate monitoring for a better yield of the region crops [8]. To make use of alarming system this project has used the Blynk apps to send notifications about the bad weather conditions to the user from the weather station.

II. LITERATURE SURVEY

Previous systems that existed are only on collection of climatic updates or transmission of these updates using ZigBee, GSM or Wi- Fi or some remote mechanism [9]. All these systems, though they measure the same parameters but they lack one common thing and that is accuracy [9]. People need accurate weather conditions of the area they live in [9]. They need to know the weather condition so that they can thrive and adapt according to it [9]. Other systems collect data and predict tomorrow's weather data just like that. No patter, no observations are made [9]. This makes the prediction error prone [9]. This method is applicable only to places where there are not so many weather fluctuations occurring in those areas i.e. it is stable throughout [9]. Since

normal prediction would fail when the outliers are more [9]. Nowadays, weather stations use heavy instruments to determine the weather of the city [9]. Through weather observation systems we are able to collect the data about humidity and temperature and in keeping with current and former data we are able to manufacture the ends up in graphical manner within the system [10]. once reviewing several articles, there are presently no papers which mention observation of the mixture of temperature, lighting and humidity in one integrated system and have actuators to change these settings [10].

III. PROPOSED SYSTEM

The proposed system consists of some extra features like GPS sensors. While GPS sensor is connected to the circuit board there is no need of changing the location settings in the code every time. Once it is connected, then wherever we go GPS sensor will send the data about that particular location to the weather station and there all the sensors get the values of temperature, humidity, CO_2, rainfall values of that particular place.
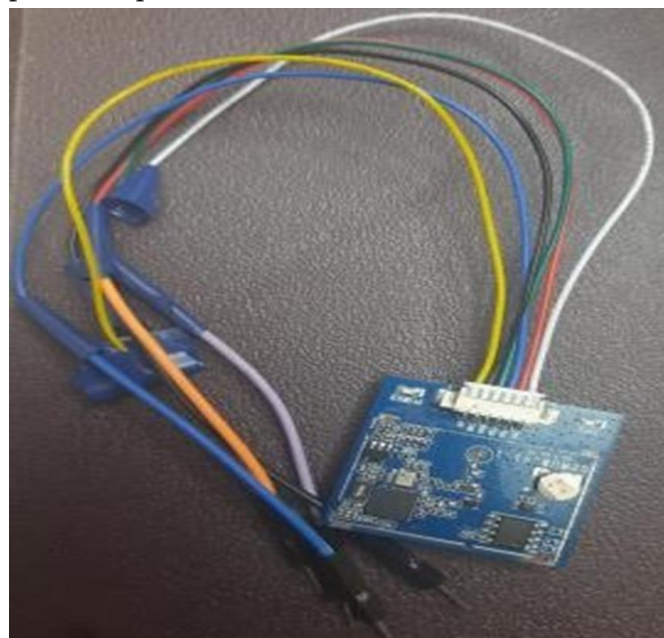
Fig.1 GPS sensor
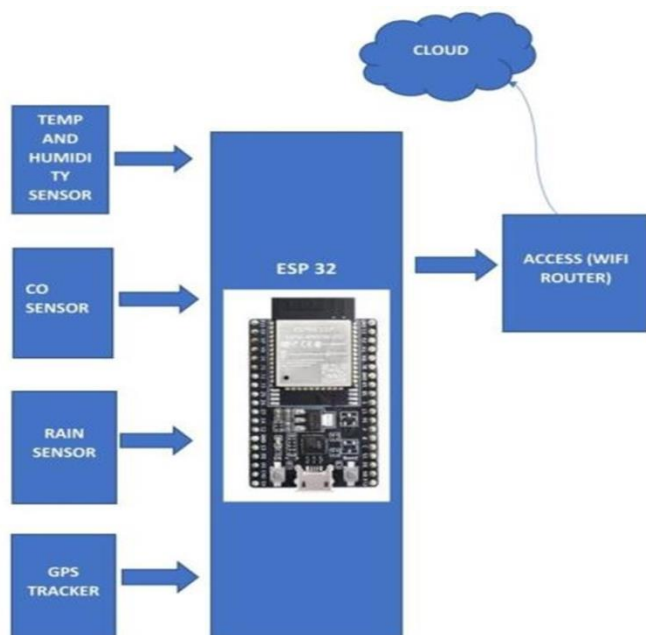
Fig.2 proposed system model

In the above block diagram, all the sensors such as temperature, humidity, CO2, rain, GPS are connected to the ESP32 microcontroller. The microcontroller consists of pins like TXD, RXD, GND, VCC etc., with which all the sensors will be connected and these hardware components will communicate with ThingSpeak using Wi-Fi hotspot connections. It also consists of LDR sensor which is used to predict whether the day is morning or night. Also, air sensor is connected which is MQ135, used for checking the value of CO_2.

These are some extra features used in this project in order to monitor the Weather efficiently. The data collected will be analyzed to configure the actual condition and the current condition by using simple formula in Equation 1.The result of this data analysis will then be used to decide the Weather state of this system to give useful information to the user about the rain and air quality condition whether is it good or bad [11].
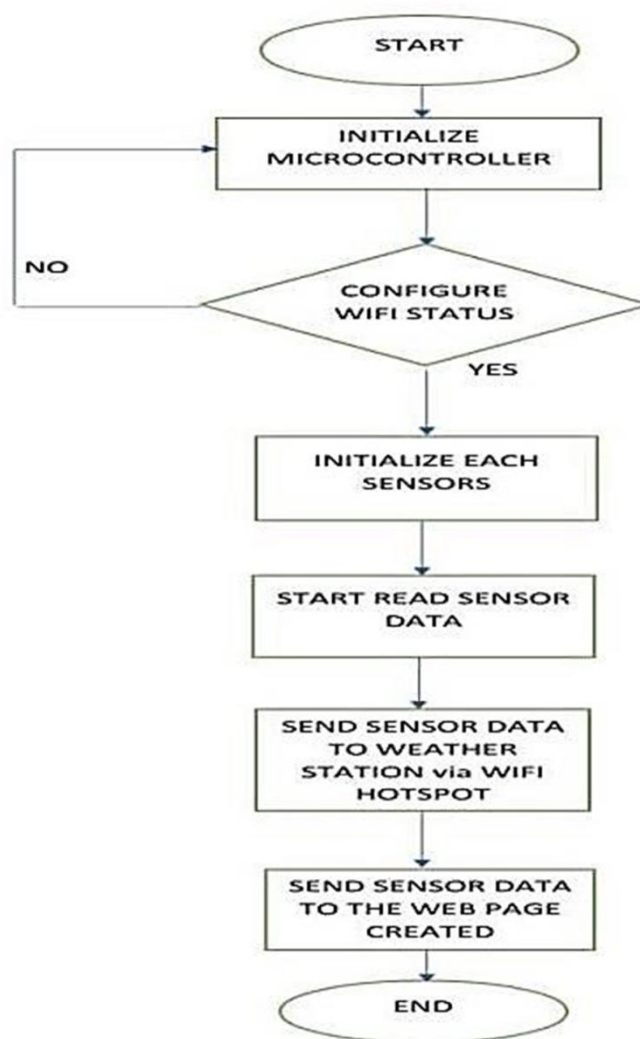
Fig.3 Flow chart

The above flowchart represents the flow of weather data from sensors to the ThingSpeak. The process of the system starts once the ESP32 configures all the sensors to read the data from the sensors [11]. The sensors that connect to the ESP32 microcontroller acts as the control unit of the system where all the data is collected here [11]. This system automatically displays the temperature, humidity, air quality, rainfall, LDR values on the specific webpage of IoT in ThingSpeak as well [11]. All these processes can be done only when the internet is configured with wi-fi or hotspot, if not it keeps on loading till it get connected to it. Internet connection of ESP32 with the ThingSpeak depends on IP network, port (COM6), baud and mainly on API key. Here serial

keyword which is used as a connection between hardware and software part.

Algorithm:

Step 1 : Create an account in ThingSpeak after creating account sign in to ThingSpeak.

Step 2 : Copy the API key present in it and paste it in the code designed in Arduino IDE.

Step 3 : Enter the Wi-Fi name and password to which the system is connected.

Step 4 : Connect to the server "api.thingspeak.com" so that there establishes a communication between the code and the ThingSpeak server.

Step 5 : Initialize the sensors DHT11, water level sensor, rain sensor, LDR sensor, CO2 sensor and finally GPS tracker.

Step 6 : If the Wi-Fi credentials matches with the system Wi-Fi, which executes two functionalities: -

1. Setup function

- Initialize the baud to 115200
- Serial connection is used to establish interaction between hardware components and software.
- Hardware components are connected to the system using data cable with which COM6 port will be enabled, which is specific to the developed code.
- once the port is enabled it starts collecting the data from all the sensors.

2. Loop function

- Firstly, it reads and collects temperature and humidity values and then it displays those in the web page.
- Next, it reads the LDR values then it compares with threshold LDR values.
- If LDR values are comparatively greater than threshold LDR values then it displays "Night" as the output else it displays "day".
- After collecting LDR value it reads MQ135 value which is a gas sensor that reads the value of CO2. It the value >= 400 ppm and value < 750ppm then it is good for health. if the value < 1200 then it

displays take care else it displays that the air is harmful to health.

- Next, it verifies whether it is raining or not using rain sensor if it is raining then it shows the rainfall value. Also, prints rain water level value in centimeters.
- Finally, it tracks the particular location and prints latitude and longitude values.

Step 7 : The above process will be done if the Wi-Fi is connected, and prints the values continuously for every 10 seconds.

Step 8 : In the other case, if the credentials of Wi-Fi are not matching it continues to wait until proper credentials are given.

IV. EXPERIMENTAL DESIGN

Hardware Experiment:

Fig. 4 ESP32

ESP32 is a series of low cost, low-power system-on-chip microcontrollers designed by Shanghai, China. It has integrated Wi-Fi and dual mode Bluetooth radios. It has many capabilities of the Arduino and also can be programmed easily with Arduino IDE software and easy to upgrade t h e path to wireless communications for Arduino users.

ESP32 cab support many peripherals such as capacitive touch, ADC, DAC, I2C, SPI, UART,I2S,

PWM. It is appropriate for IoT projects and enables Bluetooth communications for smart phone applications.

Sensors:
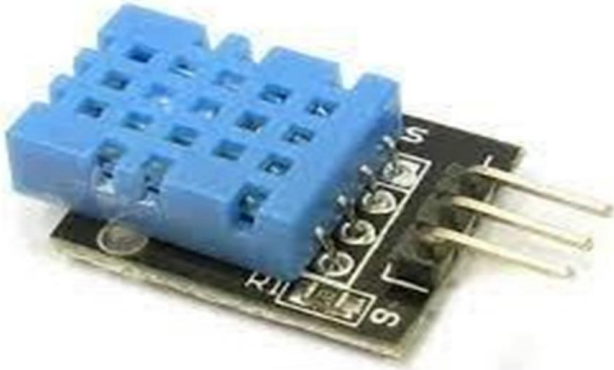
Fig.5 Temperature and humidity sensor

It utilizes a capacitive humidity sensor and a thermistor to gauge the surrounding air, and releases a digital data on the data pin (no analog information pins required) [5]. The main genuine drawback of this sensor is that you can just get new information from it for every 2 seconds, so when utilizing our library, sensor readings can be up to 2 seconds lagging. It works on 3 to 5V power supply [5]. Good for 20- 80% humidity readings with 5% accuracy and for 0-50°C temperature readings ±2°C accuracy [5].
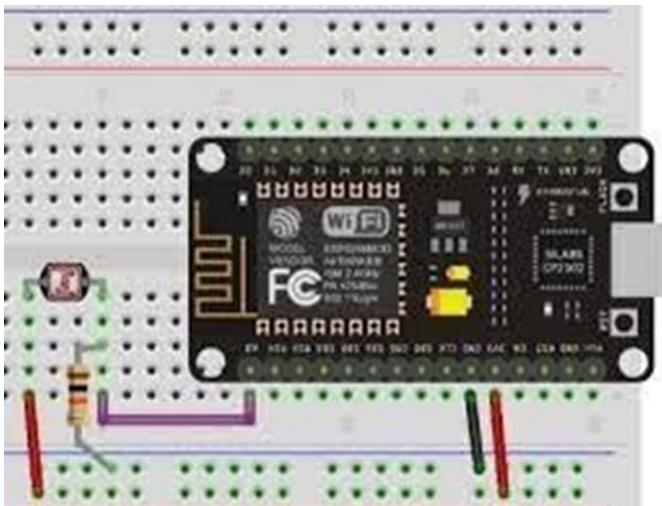
Fig.6 LDR Sensor

The sensor that can be used to detect light is an LDR. Since the LDR gives out an analog voltage, it is

connected to the analog input pin on the Arduino. The Arduino, with its built-in ADC (analog-to-digital converter), then converts the analog voltage (from 0-5V) into a digital value in the range of (0- 1023).

Fig. 7 CO_2 sensor

This sensor is used to measure air quality equipment and is suitable for detecting and measuring of NH3, NOx, Alcohol, Benzene, Smoke, CO2. The MQ-135 sensor consists of digitalpins which ensures the sensor to operate even without a microcontroller and that comes in handy when we try to detect only one particular gas.

Conversion factors 1 ppm= 1.145 mg/m³

     1 mg/mg = 0.873 ppm [5].

Software Experiment:

THINGSPEAK: According to its developers, "ThingSpeak" is an open-source Internet of Things (IOT) application and API to store and retrieve data from things using the HTTP protocol over the Internet or via a Local Area Network [5]. ThingSpeak enables the creation of sensor logging applications, location tracking applications, and a social network of things with status updates" [5]. ThingSpeak has got integrated support from the numerical computing software MATLAB from MathWorks allowing ThingSpeak users to analyze and visualize uploaded data using MATLAB without necessary of the purchase of a MATLAB license from MathWorks [5].

V. SIMULATION RESULTS

After collecting data from all the sensors, it will be sent to the ThingSpeak. Firstly, we need to create an account in ThingSpeak. After signing up there are various options like private view, public view, channel settings etc. Once private view is chosen we can see our data coming continuously once in every 10 seconds. In this project the delay given to collect each data is 10,000 msec. Delay in collecting the data can be changed according to user's requirement. Here, the web page gives the information of all variations in weather parameters in a particular region at particular date and time.
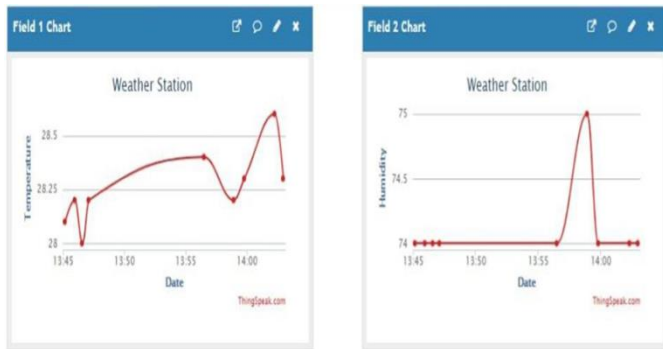
Fig 5.1 temperature and humidity values vs date and time

Fig 5.2 LDR and ppm values vs date and time
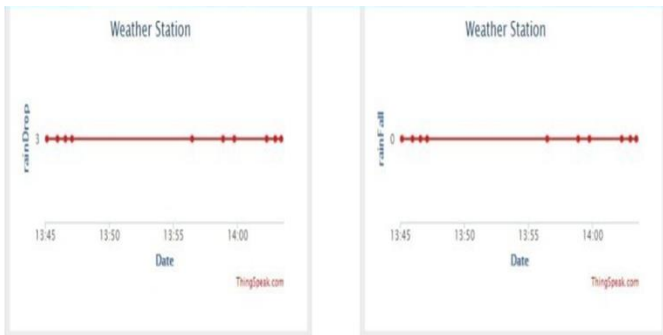
Fig 5.3 rain drop and rainfall values vs date and time
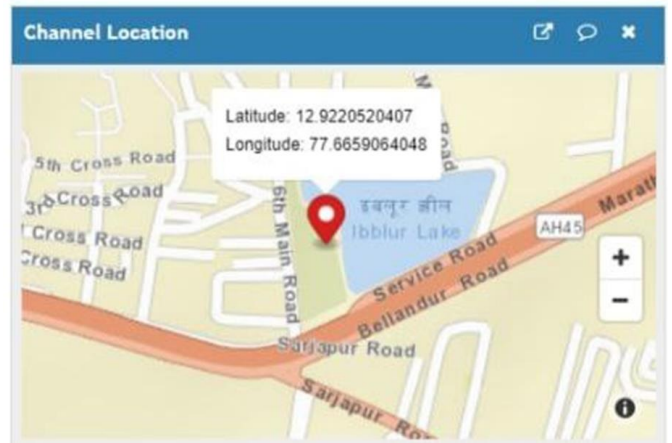
Fig 5.4 latitude and longitude of a place.

Comparative Analysis:

Below table represents the weather changes in humidity, temperature and wind speed from minute to minute.

TIME	HUMIDITY (%)	TEMPERATURE ©	WIND(MPH)
17:30	29	16	1
17:31	28	16	4
17:32	25	16	5
17:33	33	17	4
17:34	22	19	18
17:35	23	18	25

Table no 5.1 variations in weather changes from minute to minute.

The above values are observed to be less effective without having particular location. Predicting the values of temperature, humidity, air quality becomes a tedious task when there is unavailability of location. In order to avoid that we have used a GPS tracker in this paper so that monitoring of the weather condition of a particular place can be done effortlessly.

Location (lat&long) Weather	12.9598 577463 & 77.7139 310679	12.988875 338679774 & 77.728902 94602526	12.9220520 4069514 & 77.6659064 0479784

parameters			
Temperature	28	28.2	28.3
Humidity	80	74	75
LDR	4095	4095	4095
PPM	2570.62	2196.19	2305.8
Rain Drop	3	3	3
Rain Fall	0	0	0
Time	13:27:27	13:47:47	14:02:16

Table 5.2 Weather Parameters of 3 locations

The above table contains the information of weather parameters of three locations where the location is marked using latitude and longitude values. The above table contains all the weather parameters like temperature, humidity, LDR values etc. For every location weather changes slightly, so this can be represented using graphs where for every point to point there will be a change in the weather condition. Here without latitude and longitude location if we try to monitor the weather it will be difficult to find which location has which temperature or humidity or ppm values etc. So, in order to avoid those confusions, we use GPS tracker for finding exact location and monitor the weather changes.

VI. CONCLUSION

This project exhibits the design and implementation of weather monitoring system using IoT which is used to monitor the weather changes of a place. Sensors used in this project are said to be a reliable solution for forecasting the changes in the weather. GPS tracker used in this project will be helpful to track the latitude and longitude values automatically and forecast the weather changes of that place so that we can predict the weather changes and send this information to the end users which is useful to monitor health condition of the people and to reduce the pollution by taking few precautionary measures.

VII. FUTURE SCOPE

Some machine learning algorithms can be added to this project so that it will be useful to farmers for doing agriculture. They don't need to stay near the crops for a long time, instead using machine learning we can predict weather and we can check the crops sitting at home itself.

VIII. REFERENCES

[1]. V. A. S. G. M. kulkarni, "Weather Reporting System Using FPGA : A Review," ICTES, vol. 4, no. 11, pp. 319-320, 2017.

[2]. J. P. D.,. D. F. ,. p. S. Carlos Moron, "Design, Development and Implementation of a Weather Station Prototype for Renewable Energy Systems," MDPI, pp. 1-13, 2018.

[3]. K. F. a. F. A. Karim F, "Monitoring system using web of things in precision agriculture," FNC, vol. 110, pp. 402-409, 2017.

[4]. K. R. K. a. S. A, "An IoT based weather information prototype using WeMos," IC3I, no. 1, pp. 612-616, 2016.

[5]. H. H. A. G. S. P. H. P. Girija C, "Internet of Things (IOT) based Weather Monitoring System," IJERT, vol. 6, no. 13, 2018.

[6]. J. J. Joe F, "IoT Based Weather Monitoring System for Effective Analytics," IJEAT, pp. 311-315, 2019.

[7]. M. a. K. U. Nallakaruppan, "IoT based Machine Learning Techniques for Climate Predictive Analysis," IJRTE, pp. 171-175, 2019.

[8]. D. C. F. F. L. B. L. A. M. L. M. C. E. D. M. J. T. M. a. D. S.F.D. A. Monteiro M S, "University campus microclimate monitoring using IoT," ICTES, pp. 3-7, 2019.

[9]. P. Kedia, "LOCALISED WEATHER MONITORING SYSTEM," IJERGS, vol. 4, no. 2, 2016.

[10]. J. S. J. S. N. S. P. K. Yash J Joysher, "Arduino Based Weather Monitoring System," IJRET, vol. 5, no. 10, pp. 419-422, 2018.

[11]. F. P. a. M. Z. hasan, "Development of IoT Based Weather Reporting System," ICTES, 2020.

[12]. Y. S. a. S. S. Kodali R K, "Smart Farm Monitoring Using LoRa Enabled IoT," ICTES, p. 18, 2018.

[13]. G. V. K. S. a. T. A. Gahlot N, "Zigbee based weather monitoring system," IJES, vol. 4, no. 4, 2015.

V2-Vizard Visualiser

Deepen Shrestha[1], Dipesh Shrestha[1], Gaurav Shrivastava[1], Dr. S. Sridevi[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Covid-19 was declared as a pandemic by WHO on 11th March, 2020 disrupting the whole world day to day living and collapsing the economy of every nation. Wearing a mask became a vital task to prevent the transmission of the virus and is a new normal to the whole world. In the present context, it is very important for every store from bigger scale to smaller scale to enforce the rule of wearing mask which is a crucial task in order to implement. Therefore, the need of the face mask detector in the stores are must. Not only stores but it is seen to be important in each and every place This paper presents a simple approach to the solution for the stores as well as other places such as offices, homes etc. to implement the method of face mask detection using the concept of Deep Learning. The concept includes the implementation of the packages like Tensor Flow, Keras and a frontend framework Flask.

Keywords: Covid19, Deep Learning, Keras, Tensor Flow, Face Mask Detection

I. INTRODUCTION

According to the official dashboard of the Worldometer, coronavirus 2019 (COVID-19) has infected more than 159 million people and caused more than 3.3 million deaths worldwide [1]. People with COVID19 had a greater prevalence of symptoms that were reported to progress from mild manifestations to serious illness. Over time, covid-19 has already become a series of mutations of many variations.

Many strains of the SARs-COV-2 virus have been contagious locally and globally. According to the Centre for Disease Control and Prevention (CDC), three classifications of COVID-19 variants, namely Variant of Interest (VOI), Variant of Concern (VOC), and Variant of High Consequence (VOHC) B are

considered. .1.1 .7, also known as the UK exception, found in the southeast of England and currently identified as the Variant of Concern (VOC). Given the recent increase in the number of COVID cases and the spread of problems not only for the most vulnerable, but also for young people, it is an indication that the new variation of COVID poses a significant risk to human health and is associated with an increased risk of death compared to the first. With the advent of coronavirus, humans have been experiencing a wide range of symptoms. Although the list has continued to expand and pass quickly, the most common symptoms remain the same [3].

In times like epidemics, it is important to be aware of the old symptoms of COVID-19, so that the virus can be detected at an early age. Some of the most common

COVID symptoms are Fever, dry cough, sore throat, Runny and runny nose. [2].

The public should know whether to wear a resource management mask or a COVID-19 disgust. Points may be of interest to the use of masks lying down to reduce the risk of injury from a dangerous person during a "pre-symptomatic" period and the stigma of isolated people who wear masks to prevent the spread of the virus.

The WHO emphasizes the priority of medical masks and respiratory health care providers [4]. Therefore, facial recognition has become an important activity in the current global community. Finding a face mask involves finding the location of the face and deciding whether to have a mask on or not. The problem is too close to finding a common object to find categories of objects. Face identity works by distinguishing a specific group of things that means a face. It has many applications, such as private driving, education, surveillance, and so on [5].

This paper presents an easy way to achieve the above goal using basic Deep Learning (ML) packages such as Tensor Flow, Keras, and Flask. The other paper is organized as follows: Phase II discusses the functionality of the database used. Section III sets out the details of the packages used to construct the proposed model. Section IV provides an overview of our approach. The results of the evaluation and analysis are reported in section V. Phase VI concludes and focuses on the line of future activities. Literature review

Researching on different papers, we have illustrated the following researches:

1. Ben Ayed et al. suggested a way in 2015 to get text data based on video frame framing with big data analytics.
- Rotted video frames into various set size blocks and those blocks are analysed using the har wavelet transform process.
- In addition, they use the neural network to distinguish text and non-text blocks.

- However, this study needs to focus on extracting regions to address and remove noisy regions and extracting text as categories.

2. Serdar Yegulalp, Chief Writer at InfoWorld in 2016 describes TensorFlow an open library of numeracy and machine learning.
- According to him, an open source library that is compatible with Python arithmetic that makes machine learning faster and easier.
- It also defines a TensorFlow bundle together with a host of machine learning and in-depth learning (aka neural networking) models and algorithms and makes them useful in a common analogy.
- Comparison with TensorFlow uses Python to provide a simple front-end API for building framework applications, while using those applications with high C ++ performance.

3. Karanbir Singh Chahal and Katal Dey researched Modern Object Detection Literature using Deep Learning and raised a paper in 2018.
- This paper focuses on two types of object acquisition algorithms- the SSD phase for one-step detection and the Faster R-CNN two-step detection phase.
- Strategies for building portable and fast machines on low-power devices are also being considered by testing new lightweight structures in use.

II. PROPOSED SYSTEM

Our proposed system addresses the needs of small scale business client as well as client who want it for their personal use. The system can be easily installed and takes less amount of installation time. Our proposed system has to go through training phase and then only detection can be done.

Training and detection are the two phases of our project model. The database was loaded with a training model and the model was developed into a training phase. The trained model is loaded, the face

is found in the photos and video streams and the output region (ROI) is extracted. Finally, the vizard visualiser display is used and the images or faces in the video streams are classified as a mask, a mask worn improperly, without a mask.
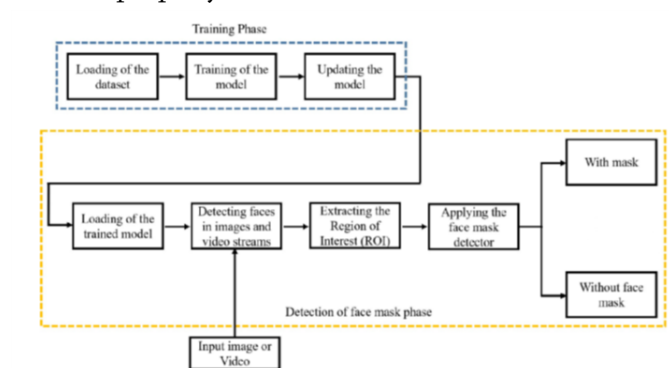
III. DATASET

Two dataset have been used for the training of the face mask detection. One dataset is used for training and other dataset is used for testing. Training dataset contains 1042 images and testing dataset contain 272 images. Fig. 1 mostly contains front face pose with single face in the frame and with different type of mask having different colours.

Fig. 1. Samples from Dataset 1 including faces with masks

The dataset contains the image of people from different ages, ethnicity and genders from different angles. Fig. 2 mostly contains front face pose with single face in the frame and without mask from different ages, ethnicity and genders.

Fig. 2. Samples from Dataset 1 including faces without masks

IV. PACKAGES INCORPORATED

A. Tensor Flow

Tensor Flow is an open source end-to-end AI segment with a wide range of hardware, libraries and network assets that allow experts to press the best of the class on the ML and its designers successfully or deploy ML-controlled applications.

Tensor Flow is available anywhere on Linux with 64-digit, MacOS, Windows, and portable subscription categories including Android and iOS. Its flexible design looks simple arrangement of calculation over an assortment of stages (CPUs, GPUs, TPUs), and from work area to bunches of workers to the versatile and edge gadgets. We are using TF-Lite convertor for the conversion of the models.[6]

1) Tensor Flow Lite converter

The Tensor Flow Lite converter adopts the Tensor Flow model and produces the Tensor Flow Lite model (a well-designed Flat Buffer format identified by the .tflite file extension). We have the following two options for using the converter:

a. Python API: This makes it easier to convert models as part of the model development pipeline, apply optimizations, add metadata and has many more features.

b. Command line: This only supports basic model conversion.

B. Keras

Keras is an API designed for people, not machines. Keras pursues the best ways to reduce psychological burden: it provides predictable and accurate APIs, limits the number of client actions required in common use cases, and provides clear and important error messages. [7] It also contains extensive documentation and developer guidelines. Kera contains a variety of squares for normal neural network construction, for example, layers, targets, work capabilities, editing agents, and a large group of apparatus that makes working with images and text details easier to optimize encoding needed for deep neural organization. The code has been simplified on GitHub, and network integration includes the GitHub problem page, and the Black channel. Cameras is a Python balanced library for in-depth discovery that can work beyond Theano or Tensor Flow. It is designed to make in-depth learning models as fast and easy as possible in testing and development. It works on Python 2.7 or 3.5 and can work consistently on GPUs and CPUs provided for basic properties. It is submitted under the consent of the patient MIT.

C. Flask

Flask is a small web framework written in Python. It is set up as a small framework because it does not require any special tools or libraries. It does not have a database extraction layer, form authentication, or other items where existing pre-existing libraries provide similar functions. However, Flask supports extensions that can add app features as if they were made to Flask itself. Extensions are available for related object maps, form verification, and upload handling, various open-source verification technologies and many framework-related tools. Flask is a popular Python web framework, which means a third-party Python library used to develop web applications. Flask is considered to be more Pythonic

than Django's web design because in most cases the Flask web application is clear. Flask is also easy to get started with as a start-up because there is a small boilerplate code for getting a simple and efficient app.

D. CNN Model
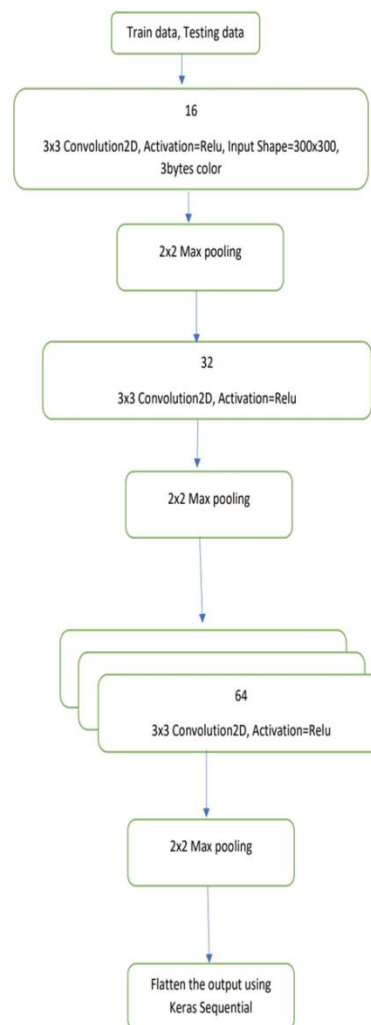
Fig. 3 CNN Architecture

Our model consists of 1 input layer which will accept the image input shape of 300X300 with the 3 bytes colour that is RGB. The images which is fed in the CNN is properly converted into that size. The activation metrics for us is Relu. We are making use of it because we are rescaling the image which contains 255 pixels into 0-1 value by using (rescale=1/255). The relu will return the value of X if X>0. Else it will return 0. The image is then passed through 3X3 convolution2D. This convolution in

CNN will help our model to extract the features from the image. In short it will isolate the features from the image. So, if we will pass the 300x300 image in the convolution it will give an output of 298x298. The image is then fed into the 2x2 max pooling. The max pooling compress the information to make it more manageable. It will directly reduce the size of the image into half. So the output from that will be 149x149. It is then fed into another convolution2d with activation relu and with 32 nerons. Then the output is fed to the 2x2 max pooling. The image output is then fed into 3 layers of 3x3 convolution2d with activation relu. It will give us the square of output. Then the final output is flattened into a 1 Dimension using Sequential provided by the Keras. Our model consist of 512 hidden layers and 1 output layer. The activation for the hidden layer is "relu" and the activation for the output layer is "sigmoid".

V. WORKING

First of all the images are collected from various sources. The images consist of people with and without mask. All the collected images are compressed into a zip file and uploaded into a cloud storage (Google drive). Then these compressed images are given as an input to that measure real-world conditions and convert samples into digital numerical values that a computer can use, we call data acquisition. Input and then split into training and testing databases on a ratio of 8:2. The training data goes under Image processing, the processed images are augmented in size, amount. Then in order to prevent neural networks from over fitting, drop out is performed. Finally feature extraction is done to where a large number of pixels of the image are efficiently represented in such a way that interesting parts of the image are captured effectively. This is then followed by creation of training model where testing data are feed into to test the accuracy of the model.

VI. RESULT AND ANALYSIS

The model is trained, validated and checked upon two datasets of people wearing mask and not wearing mask. The built model is expectedly to have an accuracy of 93%. The main reasons behind achieving this accuracy lies in Max Pooling. Here, fig (4) shows the accuracy in a line graph as well as the loss of training data loss.
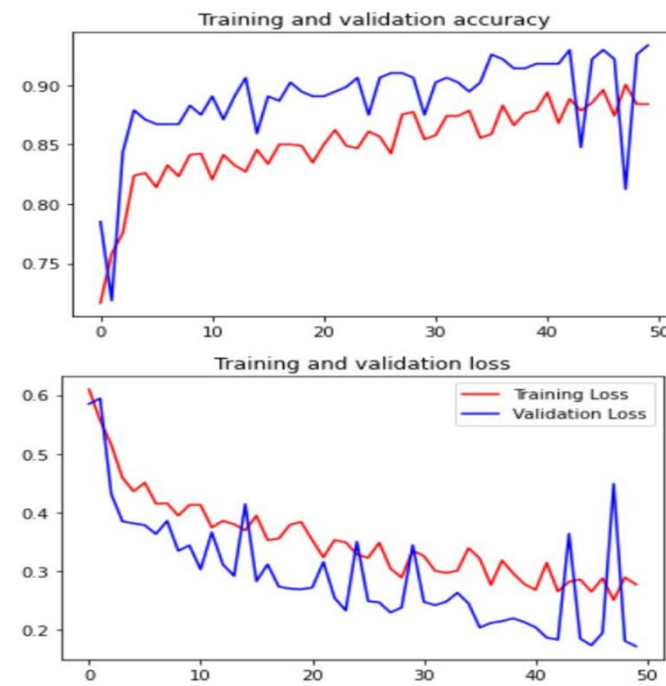
Fig. 4. Graph *showing* the accuracy of the model and training loss of the model

The dataset are trained for 50 epochs which is shown in the fig (5) giving an accuracy of training as 89% and validation accuracy of about 93% as shown in the fig(3).
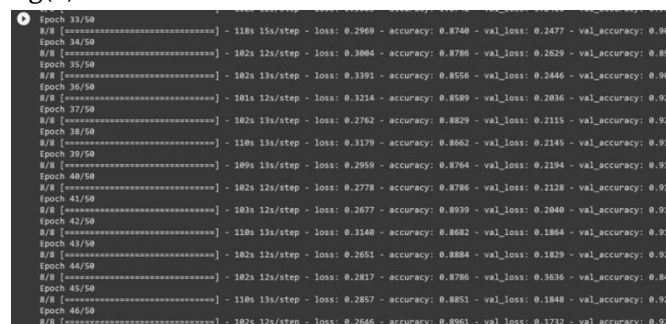
Fig. 5 Training of the epochs

With the 93% percentage of validation accuracy, some data were tested in the model and accurate result were seen. The result were as shown in the fig(6) and fig(7) validating whether a person is wearing a mask or not.
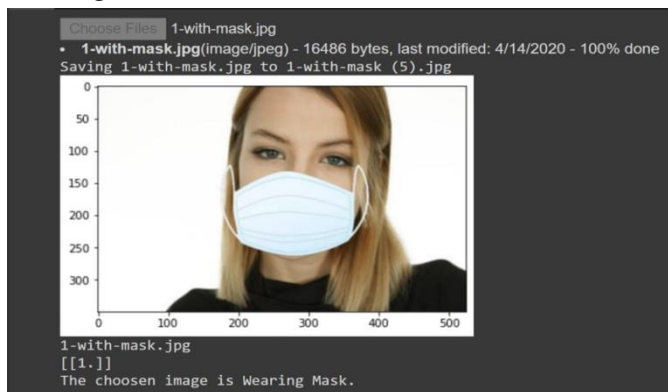
Fig. 6 Validation of person wearing mask
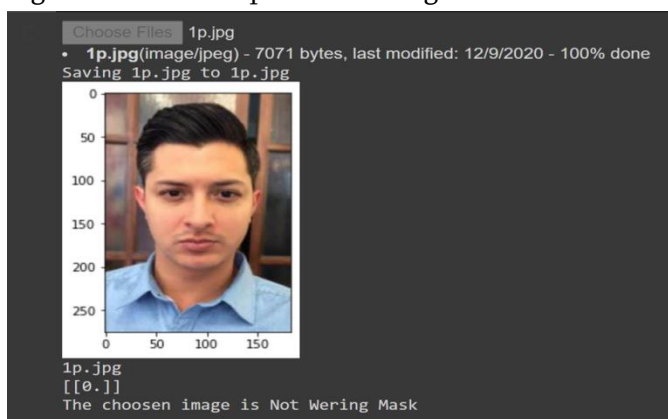
Fig. 7 Validation of person not wearing mask

VII.   CONCLUSIONS

In this paper, we briefly describe the motivation for our work and the problem the project is facing. After that, we demonstrated the learning function and performance of the model. Using Basic Learning Tools and Flask simplified techniques the method should achieve high precision detail in capturing results in the form of a web service. It can be used for a variety of applications. Wearing a mask is a must in the current situation when you look at the Covid-19 problem. Many social service providers ask customers to wear masks appropriately to access their services. The model used should contribute significantly to the public health care system. In the future it can be expanded to find out whether a person is wearing a mask properly or not. The model can be further developed to determine whether the mask is infected or not eg the type of mask is undergoing surgery, N95 or not. Also, the model can also be improved with the feature of getting the temperature, coughing and sneezing detection soon.

VIII. REFERENCES

[1]. Worldometer, Covid-19 Coronavirus Pandemic: total cases2021[Online].Available:https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?

[2]. "Coronavirus Disease 2019 (COVID-19) – Symptoms", Centres for Disease Control and Prevention,2020.[Online].Available:https://www.cdc.gov/coronavirus/2019ncov/symptomstesting/symptoms.html. 2020

[3]. Times of India, Coronavirus Mutations: How COVID Symptoms Vary In Different Strains[Online] Available:https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/coronavirus-symptoms-in-all-mutations-how-covid-symptoms-vary-in-different-strains-heres-everything-you-need-to-know/photostory/82216086.cms.

[4]. W.H.O., "Advice on the use of masks in the context of COVID-19: interim guidance", 2020.

[5]. M. Jiang, X. Fan and H. Yan, "Retina Mask: A Face Mask detector", arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/2005.03950. 2020

[6]. Alaki Sethia,Build your Own Object Detection Model using Tensor Flow API[online] Available: https://www.analyticsvidhya.com/blog/2020/04/build-your-own-object-detection-model-using-tensorflow-api/

[7]. Jaosn Brownie,Your First Deep Learning Project in Python with Keras Step-By-Step[online] Available:

https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/

[8]. Wikipedia,Flask[online]Available:https://en.wiki
pedia.org/wiki/Flask_(web_framework)

[9]. Prabhu,Understanding of Convolutional Neural Network (CNN) — Deep Learning[online].Available:
https://medium.com/@RaghavPrabhu/understan
ding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[10]. GeeksforGeeks,Understanding Google Net Model – CNNArchitecture[online].Available:https://www.geeksforgeeks.org/understanding-googlenet-model-cnn-architecture/

Vox-Mail : Voice Based Email Service For Visually Impaired

D.A Anupama[1], Litta Joseph[1], Misba Banu[1], Tinu N.S[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

The internet has become one of the most important aspects of modern life. Every human being uses the internet to gain access to knowledge and information. However, blind people have difficulty accessing these text materials, as well as using any internet-based service. The development of computer-based accessible technologies has provided many opportunities for the visually disabled all over the world. Audio input virtual environments, such as screen readers, have greatly aided Blind people in accessing internet applications. We define the Voicemail system architecture that a blind person can use to quickly and easily access e-mails. This research's contribution has helped blind people to send and receive voice-based e-Mail messages in their native language using a computer.

Keywords — Visually challenged people, IVR, Speech to text converter, Mouse click event, Screen reader.

I. INTRODUCTION

We have seen that many fields have been dramatically revolutionized by the onset of the Internet. The Internet has made people's lives so easy that people today have access to any data they want at home. Communication is one of the main fields that the Internet has revolutionized. And the first thing that comes to mind when we talk about communication over the Internet is e-mail which one of the revolutionary electronic technologies. Over about 4.1 billion or so email accounts has been created in the year 2014 and an estimated more than 5 billion accounts will be created by the end of 2018, making emails the most used and reliable medium of communication.

Visually disabled individuals cannot use the most popular mail systems we use in our everyday lives because they do not have any facilities so that the person in front can hear the contents of the screen. Since they are unable to imagine what is already viewed on the screen, they are unable to find out where to click and perform the appropriate actions. There have been different technologies available in this world today, such as screen readers, etc., but they are not as effective to perform actions. Using a computer for the first time is not as easy for a visually impaired person as it is for a regular user, even though it is user-friendly. While there are screen readers available, these individuals still face minor difficulties.

Screen reader not only reads those contents that are viewed on the screen and even the people will have to make use of all the keyboards shortcut for performing certain action because the screen will not

be able to track back the actual position of the mouse. Therefore a person who is using a computer for the first time will not be able to use the services because as it is not aware of any of the specific keys location present in it. Screen readers read the content sequentially, so that, as a user can only understand all contents present in the screen if they are in the standard HTML format. Thus the modern web page those which do not follow the model in order to make these website a user-friendly and to create more problems for these individuals.

The project aim in the development of an email that helps even the naïve, blind people and also illiterate peoples to be able to send their mails by using all the service for the communication and without any past training. This system uses IVR- Interactive voice response, allowing anyone to use only their voice to manage their mail accounts and be able to send or even read as well. This device will be able to prompt the users in performing certain actions with specific voice commands, and the user will respond to them. The major use of this method is that the usage of the keyboard is totally ignored, only by voice and mouse clicks, the user is required to respond.

II. RELATED WORK

The innovation of functional computer-based technologies has opened up many pathways across a vast majority of the globe for the visually impaired. Screen readers have greatly increased the accessibility of internet software for blind people, such as audio feedback-based virtual environments. In this paper, a design of the Voice Mail system is provided using which the visually impaired will be able to access-mails conveniently and efficiently.[1]

A search engine that solely supports the interaction between Man and Machine in the form of voice. As existing search engines receive the user's request in text format and respond by downloading and displaying the documents from the server here they have proposed a web page reader and voice-based

search engine that allows users to use only their voice to control the web browser.[2]

A research work deals with the design and implementation of the Raspberry Pi speech recognition system for critical use by visually impaired individuals. The main objective of the study is to provide a simple, affordable, user- friendly and portable device for visually impaired people to use multimedia operating system applications such as text, music player and dialing system through the GSM module interface. A low-cost Raspberry Pi board is used to execute all of the above. Therefore the aim of creating a portable system was designed using offline speech recognition at a low cost.[3]

A Voice-Mail architecture research paper allows blind individuals to access e-mail and other multimedia features. There is a GUI update to the GUI of an existing mail server on this device. This architecture also offers more functionality than the current GUI, which makes it easier for a blind individual to use the application. This device minimized the load of recording keys and typing characters, in addition to providing the mailing facility easily and effectively. It was achieved for security purposes by introducing voice-to-text, text-to-voice techniques, and using fingerprint scanning techniques to enable blind people to easily access the application.[4]

An Android app that allows blind users to access their email without the assistance of others. The techniques used to implement this application are: Speech to text conversion in android, Text to speech conversion in android, API key to access Gmail and Google map as provide web services to access mails and obtain current location of the user and a Human presence detection system To ensure privacy of sending message, a hardware (PIR sensor, LED and buzzer) to identify the presence of human.[5]

III. EXISTING SYSTEM

The standard mail service is only useful for individuals who can see and type as well. Web-based

emails designed for HTML and CSS are not created while keeping everyone's accessibility in mind. Current mail services do not provide visually impaired people with easy access because they are in written format or some kind of attached data and there is no readable information. Out of the opportunity to hear the mail sent to their mail addresses. While we have screen readers that allow these individuals to access desktop apps, we do not have any technology or device that can help these individuals access web apps. Communication is the primary objective of the use of emails. But the current mail service system fails to provide individuals with user friendliness. This indicates that the use of these types of systems has a number of drawbacks. The standard mail system can't be used by the visually impaired individuals. Owing to the lack of truly voice-based software, the visually impaired individuals are unable to communicate with web-based applications.

IV. PROPOSED SYSTEM

The proposed system is based on a completely different approach and is not like the current mail systems. Accessibility is the most critical factor that has been kept in mind while developing the proposed system. A web system is said to be perfectly usable only if all types of persons, whether normal or disabled, can use it efficiently

This usability is not supported by the existing systems. The system we are designing is therefore entirely distinct from the existing system.

Unlike the current system, which focuses more on the user- friendliness of ordinary users, our system focuses more on the user-friendliness of all categories of people, which includes ordinary people ,people with visual impairments and illiterate individuals. The entire system is based on IVR- interactive voice response. The computer will prompt the user to perform certain operations to access the respective email services. One of the major benefits of this

system is that the keyboard need not be used by the user and all operations involve mouse clicks.

Now the question arises as how can the visually impaired person find the location of the mouse pointer. As particular location cannot be tracked by the visually challenged person the system has given the user a free will to click blandly anywhere on the computer screen. The IVR will specify which type of click will perform a certain function .Thus user does not need to worry about the mouse location at all.
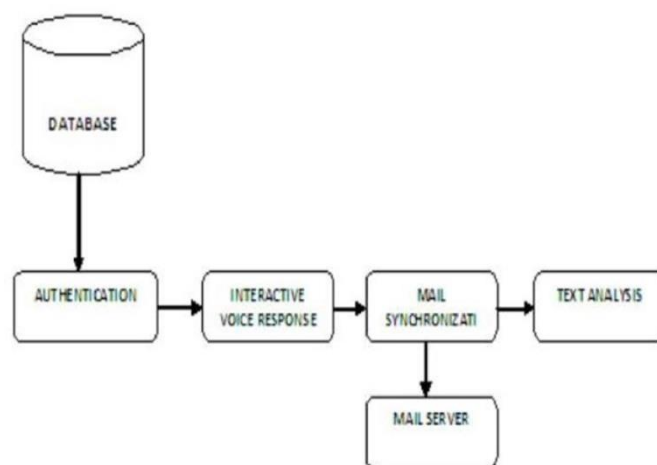
Fig. 1 System Architecture.

V. FUTURE SCOPE

There is wide scope and future work of this system as many enhancements can be done for the system which we are implementing such as including different languages, to know about the mails if they are spam mails or not, functionality of accessing the deleted mails and spam mails. Along with these functionalities, this system can be enhanced in a way such that it can also send attachments from users device which are more beneficial for people who are Visually impaired.

VI. CONCLUSION

The major project Vox-Mail: Voice Based E-mail for the Visually Impaired will help all the visually

impaired and the illiterate people to overcome all the difficulties as well as the drawbacks which they were facing while accessing the reports, documents and important files from emails. This major project will eliminate the usage of keyboard and its shortcuts which was used to traverse through all the events. Along with the elimination of keyboard usage, the usage of screen readers will also be eliminated. Thus, by reducing the load of usage and understanding all the functionalities of the keyboard.

The voiced based email can be used by visually impaired people of any age groups. Thus, this system provides many features like recognizing the user logged in through his/her speech, from speech to texts conversion and the texts to speech conversion along with the ability of speech reader which can be accessed easily without facing any difficulties. Thus, this project provides a bridge between the visually impaired so that they use the technology in a friendly manner and by providing equal opportunities.

VII. ACKNOWLEDGEMENT

VIII. REFERENCES

[1]. Tirthankar Dasgupta ,R. Ghose, & A. Basu, "Architecture of A Web Browser for Visually Handicapped People", In Students Technology Symposium (Techsym), IEEE, 2010.

[2]. Ummuhanysifa U., Nizar Banu P. K., "Voice Based Search Engine And Web Page Reader", In International Journal Of Computational Engineering Research (IJCER), 2013.

[3]. D Kiran kumar , "User Interface for Visually Impaired People" IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) Volume 12, Issue 1, Ver. II (Jan.-Feb. 2017).

[4]. Asst. Prof. Naziya Pathan, Nikita Bhoyar, Ushma Lakra, Dileshwari Lilhare , "V-Mail (Voice Based E-Mail Application)", In International Research Journal of Engineering and Technology (IRJET), Volume-06, Issue-03 (Mar 2019).

[5]. Latha L, Babu B, Sowndharya S, "Voice Based Email with Security for Visually Challenged "International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-6 S3, September 2019.

Smart Farming Using IoT: Crop Protection

P Vanajakshi[1], Harshitha G[1], Lavanya R[1], Deepika R[1], Prajay Kumar Jain H V[1]

[1]Department of Computer Science and Engineering, Vivekananda Institute of Technology, Bengaluru,

Karnataka, India

ABSTRACT

The farm lands of farmers are situated miles away from his home. Farmer sometimes need to visit is farm land to monitor and to start and stop the water supply to the field for irrigation purpose. In India, villages have scarce in availability of non-stop power which results many trips a day to operate the water pump. Farmer cannot monitor is field for 24×7 and even cannot protect the crops from natural disaster like unconditional rain. Main problem that farmers have been facing all these years and even now is power generation for irrigation where still we lack to find the permanent solution, in the existing system power generation is carried out by the conventional energy sources. Protecting the farm from heavy rainfall and also from animals attack is also a big challenge to farmers at present.

Keywords — Power Generation, Irrigation, Conventional Energy Sources, Natural Disaster

I. INTRODUCTION

The goal is to supply water to the fields by using sun based force as the primary source. In this the client can water fields and shield his harvests from anyplace by utilizing GSM strategy. The principle objective is to shield the yields from hefty precipitation and outrageous climate conditions. Producing the force from downpour water and sunlight based board. Sending data to the approved individual. Client can screen their huge fields, woodlands, gardens from anyplace, this can be utilized in agrarian fields which is extremely useful and valuable for the ranchers as sun oriented energy is bounty and sustainable, individuals can rely on this energy accordingly by saving the other non-sustainable power sources, it enhances the force utilization through water assets the executives and furthermore saving government's free auxiliary power and this demonstrates a proficient and economy method of water system and this will robotize the farming area. Helps using non-sustainable sources by utilizing sun oriented energy, there is no need of workers, it saves water and clients time. It ensures the harvest and it is minimal and minimal expense insurance for water pimps during dry invalid.

II. PURPOSE

Client can screen their enormous fields, woods, gardens from anyplace and this can be utilized in agrarian fields which is exceptionally useful and valuable for the ranchers as sun based energy is bounty and inexhaustible. Individuals can rely on this energy consequently by saving the other non-environmentally friendly power sources. It advances the force use through water assets the board and furthermore saving government's free auxiliary power

285

and this demonstrates a productive and economy method of water system which will computerize the agrarian area. Helps in using non-sustainable sources by utilizing sun based energy, there is no need of representatives, it saves water and clients time. It secures the yield and it is reduced and minimal expense insurance for water siphons during dry invalid.

III. LITERATURE SURVEY

The literature review in this work made by various analysts and their methodology and conclusion are considered.

A Survey Paper on Web of things based Application Brilliant Farming Framework , 2017

[1] In this paper essentially we are zeroing in on the past existing methodology. As we can see some past existing innovation which depends on GPRS innovation. In this paper we attempt to learn IoT framework. According to the current framework, there is need of those kind of framework, which is continuous information analyzer which can send information all throughout the planet utilizing the idea of IoT. In future we can deal with Wi-Fi, Lora based innovation. In this space we can chip away at IOT based framework. Here we can give portable application office which is straightforward We can chip away at Wi-Fi based innovation so it's exceptionally modest framework. May be this framework won't confront the availability issue. Here we can likewise utilize robbery control framework. Here we will attempt to discover water level by crop type.

An Overview on the Job of IoT in Agribusiness for the Execution of Keen Cultivating, 2019

In this paper, a microwave non-obtrusive blood glucose checking framework working at around 1.4 GHz is planned and its exhibition as far as exactness and repeatability is assessed by a clinical preliminary

including 24 human subjects, with and without diabetes. [2] Direct correlation with the most precise seat top glucose analyzer shows the extraordinary exactness and repeatability of the proposed framework. Diabetes addresses one of the biggest wellbeing worries of the 21st century. It influences 7% of the total populace and is on the quick ascent, being assessed to influence 1 of every 10 grown-ups by 2040 . The danger of diabetes-related complexities can be essentially decreased by keeping up close power over blood glucose. At present numerous patients make standard estimations of their blood glucose. This typically includes pulling out a blood test by means of an obtrusive finger-prick method. This is both agonizing and awkward, bringing about diminished patient consistence and less compelling glycemic control, or by other obtrusive strategies, for example, persistent glucose observing frameworks (CGMs), with related danger of contamination. Subsequently, there is an always developing interest for a dependable non-intrusive glucose checking framework.

Creature Characterization Framework Dependent on Picture Preparing and Backing Vector Machine, 2016

When contrasted and execution of individual descriptors, acknowledgment of positive and negative picture independently was not fruitful. [3] Be that as it may, when not many descriptors joined together, exceptionally reasonable result was clear at this point exactness differed around 80%. According to the diagram which addressed the consolidated descriptor results similarly, victories were obvious which perceive both positive and negative pictures genuinely in equivalent way. The blend consequences of locale shape, edge histogram and shading structure had exceptionally low division among positive and negative segments. The blend consequences of district shape, edge histogram and shading design were given normal result contrasted and the above mix. The mix of locale shape, edge histogram, shading format and shading structure, however it perceives a decent lot of

positive pictures, neglects to isolate negative pictures in an extraordinary way. So mid setup of both informational indexes, which is mix of district shape, edge histogram and shading structure, if great outcomes in acknowledgment. In an outline contrasted and singular descriptor exhibitions, joined descriptors gave a lot of precise result.

Security of Harvests from Wild Creatures Utilizing Insightful Observation Framework, 2018

The issue of yield vandalization by wild creatures has become a significant social issue in the current time. [4] It requires earnest consideration and a powerful arrangement. Accordingly this undertaking conveys an incredible social importance as it expects to address this issue. Thus we have planned a savvy installed farmland security and reconnaissance based framework which is minimal expense, and furthermore devours less energy. The fundamental point is to forestall the deficiency of yields and to shield the territory from interlopers and wild creatures which represent a significant danger to the agrarian territories. Such a framework will be useful to the ranchers in ensuring their plantations and fields and save them from huge monetary misfortunes and furthermore saves them from inefficient endeavors that they suffer for the assurance of their fields. This framework will likewise assist them with bettering harvest yields accordingly prompting their financial prosperity.

IoT Applications in Horticulture: An Efficient Writing Audit, 2019

[5] IoT innovations permit, in addition to other things, to get data on environment, stickiness, temperature, soil ripeness to productively do distant observing of yields. On account of these advancements, ranchers can know the situation with their harvest whenever and from any spot. Then again, remote sensor networks permit controlling the states of the ranch, just as computerize various cycles. For instance, a portion of the examinations investigated in this work

utilize remote cameras to know the situation with the harvest continuously. Different examinations have utilized robots to help the assignments of exactness agribusiness, just as cell phones to keep ranchers educated about the current states of their development. Probably the most extraordinary advances that are joined with IoT to foster farming arrangements are remote sensor organizations, distributed computing, middleware frameworks, and portable applications [48, 49]. IoT innovations are now a fundamental piece of taking care of various issues in the agrarian setting. Therefore, in this work, we played out a writing survey meaning to distinguish the primary IoT applications in agribusiness, IoT-based programming and gadgets utilized in horticulture, just as the advantages given by this sort of advances. Albeit the subjects tended to in this work are vital for all individuals associated with horticulture, it is critical to make reference to the requirement for breaking down more examination works completed around the goal of natural issues intending to accomplish a manageable development of food.

Wild-Animal Acknowledgment in Horticulture Ranches Utilizing W-COHOG for Agro-Security, 2017

In this paper, we proposed another calculation for creature acknowledgment. [6] This strategy accomplished better precision on two benchmark datasets contrasted with other existing calculations. The tests were directed on static pictures. In future, we will carry out a strategy for perceiving creatures in the video. Later on, we work on improving the calculation to improve the precision.

Sensor Based Programmed Water system Framework and Soil pH Recognition utilizing Picture Handling, 2017

[7] The prior framework burns-through more water and came about to a Water Wastage. Introducing the programmed water system framework and deciding

the pH esteem it saves time and guarantees wise use of water and ranchers become more acquainted with before that what yields can be filled in his field. This framework works in regions where there is no standard inventory of power. The framework is decreasing human mediation in this manner less energy of the rancher is required. In future the modules like Computerized reasoning can be added to consequently become familiar with the example of watering the yields.

A Savvy FARMLAND Utilizing RASPBERRY PI Yield Anticipation AND ANIMAL Interruption Identification Framework, 2018

The issue of harming crops by wild creatures has become a significant social issue in the current time. [8] It requires critical consideration and a powerful arrangement. Along these lines this undertaking conveys an incredible social importance as it expects to address this issue. The proposed framework dependent on Raspberry pi is discovered to be more reduced, easy to use and less unpredictable, which can promptly be utilized to perform. A few monotonous and dull undertakings. In this undertaking the interaction is completely robotized and it doesn't make any hurt creature during repellent. Future extension in this undertaking is to identify the area of the creatures by utilizing RFID injector and GPS.

A Savvy IoT-Based Framework Plan for Controlling and Observing Nursery Temperature, 2017

To defeat the extremely prohibitive climatic conditions in the KSA, a profoundly adaptable wise framework that screens the nursery climate, produces the reference temperature, and controls the inner temperature was created. [9] The utilization of a PN model permits us to screen the nursery climate, to produce reasonable reference temperatures, and to oversee the entire framework. A controlled canopy that lessens the impacts of the Sun beams was additionally presented. The proposed framework is

self-sufficiently ready to: screen the external temperature, screen the energy utilization times of heavy traffic, screen the points of the Sun beams, produce the appropriate temperature, send this temperature as a kind of perspective sign for temperature guideline, ensure that the encompassing nursery temperature arrives at this reference temperature, lastly, goes into reserve state without assignments to achieve. The fundamental creative mark of this work is scalable design of a supporting management information system

An IoT Based Climate Checking Framework, 2020

This Climate Observing Framework is produced using minimal expense segments that are effectively accessible and can be utilized to screen a few natural boundaries. [10] This framework can be effectively be adjusted for both indoor or open air use. The proposed framework has been tried a few times with various boundaries, and have been fruitful all through. To wrap things up, this gadget can associate with the passage by means of Bluetooth, Infrared or WiFi absent a lot of configuration changes along these lines making it appropriate for various situations. This framework is consequently adaptable and versatile. In future the exploration work is expected to present a few AI methods that will give more knowledge to the client. Additionally, to oversee changes productively, the records can be kept in a protected permanent advanced record utilizing innovation like Blockchain.

IV. METHODOLOGY

In case of heavy rainfall the farmer will send a signal or a message to start the operations. Hence the crop is covered by the double coated polythene sheet over the agriculture land & crop is protected. The required protection is fabricated by four adjustable poles which enables the adjustment of height. The microcontroller is used to control this operation using IOT technology which enables the farmer to control the operation from the remote place. Even System works in

automated mode i.e. when farmer doesn't respond to the request from Wi-Fi, it checks the moisture content of the soil using moisture sensor and initiates appropriate action required to protect crop. Detection of rain using a rain sensor. Detection of soil moisture level. Generated power from solar panel is stored in battery to pump the water back to the crops.

Working:

At the point when the force supply is turned on, the GSM modem gets introduced. The GSM modem speaks with the RASPBERRY PI board utilizing AT orders. The LCD show is associated with the ADC pins of the ARM processor, to show the message. Right off the bat the processor checks for the accessibility of the sun based energy with the assistance of LDR, the sun powered board is interfaced with the stepper engine, which thus is associated with the stepper driver.

Architecture diagram:
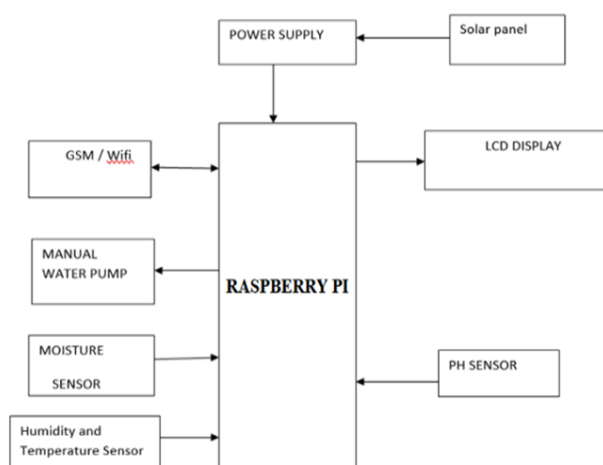
The dampness sensor checks for the dirt dampness content whose most extreme edge is kept at $1000^{\circ}c$ and least of $300^{\circ}c$. At the point when the dirt dampness content is under $300^{\circ}c$ the siphon engine will siphon the water to the recorded. The temperature sensor will quantify the encompassing temperature of the homestead.

The downpour sensor will detect the substantial downpour and shuts the board to secure the harvest. All the above data will be educated to the client utilizing GSM innovation. At the point when the force supply is turned on the RASPBERRY PI microcontroller and the GSM modem/GPRS is introduced. After the instatement, the framework ask the client either to choose programmed mode or the manual mode. When the auto mode is chosen, initially the processor checks for the accessibility of the sun oriented energy with the assistance of LDR which is utilized for detecting the daylight.

The temperature sensor detects the encompassing temperature of the homestead. At the point when it begins coming down, the siphon engine will quit siphoning the water to the field and updates the client utilizing GSM/GPRS method. When there is an unequivocal downpour the boards gave will be shut consequently to ensure the yield. An Alpha numeric LCD is utilized to show the data. When the manual mode is chosen, the data about the homestead field will be refreshed to the client just when he calls to a confirmed number given.

V. CONCLUSIONS

The venture is accordingly done utilizing RASPERRY-PI center with the assistance of GSM innovations. This task discovers application in home grown agrarian field. In regular citizen space, this can be utilized to guarantee steadfast water system of homestead field, since we have the choice of discovering dampness level of soil in a specific area[6]. The mechanized water system framework carried out was discovered to be attainable and practical for streamlining water assets for rural creation. This water system framework permits development in places with water shortage along these lines improving maintainability. The rancher can observing from any spot, and furthermore can secure his harvest. The client can send a SMS message from anyplace on the planet to work this gear. The security highlight in

the product will ensure that it works just with pre-relegated telephone numbers.

VI. REFERENCES

[1]. Y. Erdem, L. Arin, T. Erdem, S. Polat, M. Deveci, H. Okursoy and H. T. Gültas, Crop water stress index for assessing irrigation scheduling of drip irrigated broccoli (Brassica oleracea L. var. italica), Agriculture. Water Manage., vol. 98, no. 1, Dec. 2010, pp. 148–156.

[2]. S. L. Davis and M. D. Dukes, Irrigation scheduling performance by evapotranspiration-based controllers, Agriculture. Water Manage., vol. 98, no. 1, Dec. 2010, pp. 19–28.

[3]. X. Wang, W. Yang, A. Wheaton, N. Cooley and B. Moran, Efficient registration of optical and IR images for automatic plant water stress assessment, Computer. Electron. Agriculture., vol. 74, no. 2, Nov. 2010, pp. 230–237.

[4]. K. W. Migliaccio, B. Schaffer, J. H. Crane and F. S. Davies, Plant response to evapotranspiration and soil water sensor irrigation scheduling methods for papaya production in south Florida, Agriculture. Water Manage., vol. 97, no. 10, Oct. 2010, pp. 1452–1460.

[5]. S. A. O'Shaughnessy and S. R. Evett, Canopy temperature based system effectively schedules and controls center pivot irrigation of cotton, Agriculture. Water Manage., vol. 97, no. 9, Apr. 2010, pp. 1310–1316.

[6]. W. A. Jury and H. J. Vaux, The emerging global water crisis: Managing scarcity and conflict between water users, Adv. Agronomy, vol. 95, Sep. 2007, pp. 1–76.

[7]. K. S. Nemali and M. W. Van Iersel, An automated system for controlling drought stress and irrigation in potted plants, Sci. Horticult., vol. 110, no. 3, Nov. 2006, pp. 292–297.

[8]. G. Yuan, Y. Luo, X. Sun and D. Tang, Evaluation of a crop water stress index for detecting water stress in winter wheat in the North China Plain, Agriculture. Water Manage vol. 64, no. 1,Jan. 2004, pp. 29–40.

[9]. R. G. Allen, L. S. Pereira, D. Raesand M. Smith, -Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage , Italy: FAO, 1998, pp.30-55.

[10]. S. B. Idso, R. D. Jackson, P. J. Pinter, Jr., R. J. Reginato and J. L. Hatfield, Normalizing the stress- degree-day parameter for environmental variability," Agriculture. Meteorology vol. 24, Jan. 1981, pp.

[11]. Mohit kumar Navinay, Rahul Gedam, A Review Paper on Internet of things based Application Smart Agricultural System, April 2017

[12]. MUHAMMAD SHOAIB FAROOQ 1 , (Member, IEEE), SHAMYLA RIAZ 1 , ADNAN ABID 1 , (Member, IEEE), KAMRAN ABID2 , AND MUHAMMAD AZHAR NAEEM2, A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming, October 25, 2019

[13]. A. W. D. Udaya Shalika, Lasantha Seneviratne, Animal Classification System Based on Image Processing & Support Vector Machine, 15 January 2016

[14]. Vikas Bavane , Arti Raut , Swapnil Sonune , Prof. A.P.Bawane, Dr.P.M.Jawandhiya, Protection of Crops from Wild Animals Using Intelligent Surveillance System, Published in: Special Issue National Conference "CONVERGENCE 2018", 09th April 2018

[15]. Raquel Gómez-Chabla1(&) , Karina Real-Avilés , César Morán , Paola Grijalva1 , and Tanya Recalde, IoT Applications in Agriculture: A Systematic Literature Review, Published in: Springer Nature Switzerland AG 2019

[16]. Nagaraju Andavarapu1 and Valli Kumari Vatsavayi, Wild-Animal Recognition in Agriculture Farms Using W-COHOG for Agro-Security, Published in: International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 9 (2017)

[17]. Sanjay Kumawat, Mayur Bhamare, Apurva Nagare , Ashwini Kapadnis, Sensor Based Automatic Irrigation System and Soil pH Detection using Image Processing., Apr -2017

[18]. S. Santhiya, Y. Dhamodharan, N E. Kavi Priya, C S. Santhosh, M.Surekha, A SMART FARMLAND USING RASPBERRY PI CROP PREVENTION AND ANIMAL INTRUSION DETECTION SYSTEM, Mar-2018

[19]. Ahmad F Subahi , Kheir Eddine Bouazza, An Intelligent IoT-Based System Design for Controlling and Monitoring Greenhouse Temperature, 2017

[20]. Mosfiqun Nahid Hassan , Mohammed Rezwanul Islam, Fahad Faisal, Farida Habib Semantha, Abdul Hasib Siddique , Mehedi Hasan, An IoT Based Environment Monitoring System, Published by: 3 rd International Conference on Intelligent Sustainable Systems ICISS 2020

A Study of Current Scenario of Cyber Security Attacks and Solutions

Mrs. Aswathy Mohan[1]

[1]PG Student, M.E CSE with Specialization in Networks, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

ABSTRACT

Cyber Security plays a giant role among the sphere of information technology varied Governments and corporation's unit of activity taking many measures therefore on forestall these cybercrimes. This paper among the foremost focuses on challenges featured by cyber security on the foremost recent technologies. It additionally focuses on latest regarding the cyber security techniques, ethics, and place along the trends dynamic the face of cyber security. Potential choices of wise solutions to transgression are proverbial. With the technological evolution comes the progress of transgression, that typically develops new attack kinds, tools, and techniques that modification attackers to penetrate several delicate or well-controlled environments and switch out inflated hurt and even keep untraceable. supported the results of the analysis, the article presents countermeasures that firms may undertake therefore on guarantee improved security which can support in defensive their business from attackers from Associate in Nursing info security perspective. Securing the data became one altogether the foremost vital challenges among the gift day. Cyber-attacks use malicious code to vary code, logic, or data, resulting in unquiet consequences which may compromise info and cause cybercrimes, like info and fraud. Cyber-attack is additionally spoken as a network attack (CNA). throughout this paper, we have got a bent to tend to tend to mean that some examples and choices of recent cyber-attacks and describe phases of them. Finally, we have got a bent to tend to tend to conclude that alone the thought of intelligence can defend these cyber threats. The broad objective of this study is Associate in Nursing attack, threat, and vulnerabilities of cyber infrastructure, that embrace hardware and package systems, networks, enterprise networks, intranets, and its use of the cyber intrusions. The paper additionally includes a whole description and definition of cyber security, the role it plays in network intrusion and cyber acknowledge thieving.

Keywords--- Malware, Phishing, Brute force, Hijacking, Ransomware, Attacks

I. INTRODUCTION

Today man is ready to send and receive any style of information may even be AN e-mail or AN audio or video simply by the clicking of a button, firmly his information id being transmitted or sent to the choice person safely with none run of information. it's because of cyber security. Nowadays quite sixty you take care of total business transactions are done on-line, so this field needed a primary quality of security for clear and best transactions. therefore, cyber security has become a recent issue. The scope of cyber security is not simply restricted to securing the data in IT trade however else to varied completely

different fields like cyber house etc. Even the foremost recent technologies like cloud computing, mobile computing, E-commerce, internet banking etc else desires high level of security. Since these technologies hold some vital info concerning a personal their security has become a requirement issue. creating web safer and protective internet users has become integral to the event of recent services still as governmental policy.

The fight against crime wishes a comprehensive and a safer approach. it is crucial that social control agencies are allowed to research and prosecute crime effectively. today many nations and governments ar imposing strict laws on cyber securities therefore on stop the loss of some important information. In today's information-age, AN organization's dependence on network is popping into and a lot of and a lot of important facet of structure security. As utterly completely different organizations infrastructure ar interconnected in network, the extent of risk to national security has inflated dramatically. The threat to cyber security is growing. according to Kaspersky world IT Risk Report 2016, the very best causes for the most expensive info breaches are supported recent attacks that are evolving over time, that are among the subsequent order:

- Viruses, malware, and Trojans
- Lack of diligence and primitive workers
- Phishing and social engineering
- Targeted attack
- Crypto and ransomware

Everything may begin with a phishing email that uses social engineering to guide the worker to click on a link which can transfer an outbreak, malware, or Trojan. the first and most vital attribute is that the attacker encompasses a specific target in mind once he/she/they starts to form an inspiration of attack. throughout this primary section, the bad person pays several your time and resources to perform public intelligence to urge the desired info to hold out the attack. The motivation behind this attack stealing info.

Another attribute for this sort of attack the quantity of your time that they maintain persistent access to the target's network. The intent is to continue moving laterally across the network, compromising whole totally fully totally different systems until the goal is reached. to the target's network. The intent is to continue moving laterally across the network, compromising entirely totally different systems until the goal is reached.

one among the most effective challenges throughout this area is to identify the offender once they are already among the networks. the quality detection systems like intrusion detection systems (IDS) will not be enough to alert on suspicious activity happening, significantly once the traffic is encrypted. Crypto and ransomware area unit rising and growing threats that area unit creating a completely new level of challenge for organizations and cybersecurity professionals. In could 2017, the world was shocked by the foremost vital ransomware attack in history, said as WannaCry. This ransomware exploited a illustrious Windows SMBv1 vulnerability that had a patch free in March 2017.The attackers used AN exploit said as Eternal Blue that was free in Gregorian calendar month 2017, by a hacking cluster said because the Shadow Brokers. in line with MalwareTech this ransomware infected over four hundred,000 machines across the globe.

II. CYBERSECURITY

These cyberattacks unit typically aimed toward accessing, changing, or destroying sensitive information; extorting cash from users; or interrupting ancient business processes. Implementing effective cybersecurity measures is extremely arduous late as a result of their unit many devices than individuals, and attackers have gotten many innovative. In today's connected world, everybody blessings from advanced cyberdefense programs. At a private level, a cybersecurity attack might end up in everything from fraud to extortion

makes Associate in Nursing endeavour, to the loss of necessary knowledge like family photos. everybody depends on necessary infrastructure like power plants, hospitals, and money service companies. Securing these and each one all entirely totally all completely different organizations is extremely necessary to keeping our society functioning.

everybody in addition blessings from the work of cyberthreat researchers, rather a little amount a small amount just like the team of 250 threat researchers at Talos, World Health Organization investigate new and rising threats and cyber-attack ways in that within that. They reveal new vulnerabilities, educate the last word public on the importance of cybersecurity, and strengthen open supply tools. Their work makes Infobahn safer for everybody.

stage is pop into many and far of serious owing to the exaggerated reliance on laptop computer systems, Infobahn and wireless network standards like Bluetooth and Wi-Fi, and since of the enlargement of "smart" devices, beside smartphones, televisions, and in addition the various devices that represent the "Internet of things". owing to its quality, each in terms of politics and technology, cybersecurity is in addition one altogether the foremost challenges at intervals the stylish times. The New vogue period 1967 session organized by Willis Ware at the Spring Joint laptop computer Conference, and in addition the later publication of the Ware Report, were foundational moments at intervals the history of stage of laptop computer security.

Cyber security is that the appliance of technologies, processes, and controls to safeguard systems, networks, programs, devices, and knowledge from cyber-attacks. It aims to cut back the prospect of cyber-attacks and defend against the unauthorised exploitation of systems, networks, and technologies. It aims to cut back the prospect of cyber-attacks and defend against the unauthorised exploitation of systems, networks, and technologies. A 1977 organisation publication introduced the "CIA triad" of Confidentiality, Integrity, and convenience as a transparent and

simple owing to describe key security goals. whereas still relevant, extra elaborate frameworks have since been projected. Cyber security focuses on protective laptop computer systems from unauthorised access or being otherwise broken or created inaccessible. The GDPR (General information Protection Regulation) and DPA (Data Protection Act) 2018 wish organisations to implement acceptable security measures to shield personal information. Otherwise, there's a risk of sizeable fines. Cyber security is additionally a necessary business issue for each organisation.Cyber-attacks worth organisations billions of pounds and will cause severe injury. compact organisations stand to lose sensitive information, and face fines and reputational injury.

Global cyber security market share, by application, 2020 (%)
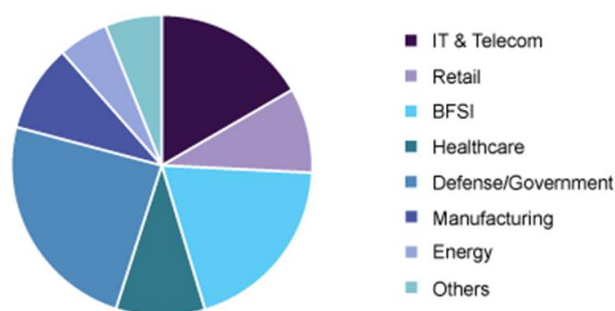
III. CYBERCRIME

Cybercrime can be a term for any criminal activity that uses a laptop computer as its primary suggests that of commission and thieving. The U.S. Department of Justice expands the definition of crime to include any criminal activity that uses a laptop computer for the storage of proof. The growing list of cybercrimes includes crimes that area unit created potential by computers, like network intrusions and thus the dissemination of laptop computer viruses, more as computer-based variations.

The findings in an exceedingly} very report discharged last year by the center for Strategic and International Studies (CSIS), —In the Crossfire:

crucial Infrastructure among the Age of Cyber war‖. supported a survey of 600 IT security managers from crucial infrastructure organizations, the report found that thirty seventh believed the vulnerability of the planet they worked accumulated over the year previous, and a couple of fifths expect a significant security incident within their sector within the next year. alone fifth of respondents to the survey believe their sector to be safe from serious cyber-attack among future five years [7]. Around 100% to twenty of the 100+ incidents recorded in BCIT's Industrial Security Incident info (ISID) to the present purpose area unit targeted attacks. The knowledgeable executive director is that the largest threat and vie a vicinity in an exceedingly} very standing case in Queensland   Australia, in Gregorian calendar month 2 hundred.

ChiChao range 71|metallic component 71|metallic element metalin his paper explores the increasing number of crime cases in Taiwan and examines the demographic characteristics of those in command of this criminal activity. As per the information purpose eighty-one.1% were male; 45.5% had some high school; sixty 3.1% acted independently; twenty 3.7% were presently listed students; and twenty 9.1% were among the 18-23 individuals, that was the majority cluster. For those listed student crime suspects, the findings show that the proportion of faculty and high school student suspects recognized sixty 9.0% (2002), 76.1% (2003) and sixty 2.7% (2004) of crime suspects in their varied years. The high rate shows that the quantity of presently listed students suspected of involvement in crime is cause for concern in an exceedingly} very survey of 100 UK of nice United Kingdom and European country organizations conducted by Activity information Management eighty 3 of the respondents believe that they are below increasing risk of Cyber Attack [Figure 1]. it has been found that cash the} and IT Sectors have money and IT sector has spare investment in cyber security as compared to sectors like central government, telecoms, and world.

2020 VICTIMS BY AGE GROUP

Victims		
Age Range[7]	Total Count	Total Loss
Under 20	23,186	$70,980,763
20 - 29	70,791	$197,402,240
30 - 39	88,364	$492,176,845
40 - 49	91,568	$717,161,726
50 - 59	85,967	$847,948,101
Over 60	105,301	$966,062,236

Ponemon Institute has been conducted a survey in June 2011 to review but well the organizations were responding to threats against network security. in an exceedingly} very survey it fully was found that organizations unit experiencing multiple created attacks against their networks [Figure 2]. fifty 9 of respondents aforesaid that their organization's network security had been successfully broken a minimum of doubly over the past year. per the findings, the common value of one data breach for U.S. organizations collaborating throughout this study was $7.2 million whereas the common value of one cyber-attack was $6.4 million.

The report of the laptop Security Institute (CSI/FBI) (Gordon, Martin, William, & Richardson, 2004) states that as regards to sixty-six of all cyber-security breach incidents, among the 280 organizations UN agency well-versed the survey, were conducted from among the organization by licenced users. additionally, AN awful seventy 2 of organizations reportable that they have no policy insurance to help them manage cyber-security risks.

Distribution of the key COVID-19 inflicted cyberthreats based on member countries' feedback

IV. TYPES OF CYBER ATTACKS

- Network security attacks
- Wireless security attacks
- Malware attacks
- Social engineering attacks

A. Phishing and spear phishing attacks

Phishing attack is that the follow of inflicting emails that appear to be from sure sources with the goal of gaining personal information or influencing users to undertake to a minimum of 1 issue. It combines social engineering and technical trickery. it'd involve associate attachment to associate email that a full bunch malware onto your digital computer. it'd even be a link to associate illegitimate information science system which can trick you into downloading malware or delivering your personal information.

Spear phishing may be targeted quite phishing activity. Attackers take the time to conduct analysis into targets and build messages that unit personal and relevant.

The solutions to avoid phishing are:

1) Critical thinking: Do not accept that associate email is that the important deal just because you

are busy or stressed. So think and read each mail carefully.

Apply necessary thinking to decipher the uniform resource surveyor.

2) Analysing email header: The attributes in reply should end in fixed domain and represented at fixed intervals.

3) Sandboxing: You will be ready to take a glance at email content throughout very sandbox atmosphere and during each interval click the mail links.

B. Ransomware

Ransomware may be a spread of malicious package, in addition referred to as malware. It encrypts a victim's information until the bad person is paid a planned ransom. Typically, the bad person demands payment throughout very sort of cryptocurrency like bitcoin. only then will the bad person send a cryptography key to unleash the victim's information. The user had to pay to "PC machine Corporation" to repair a tool. Ransomware attack has evolved long back and still growing.This is a main cyber threat. a recent attack on University of town had forced to pay ransom in bitcoins. In another ransomware attack, not only files. Exponential growth unit of measurement planning to be witnessed at intervals the ransomware attack.

Rise in Ransomware Attacks
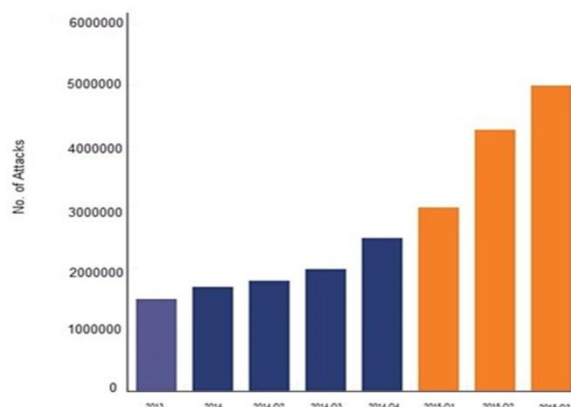
C. Malware

Malware may be a spread of package designed to comprehend unauthorized access or to cause injury to a negotiable portable computer. Malware is intrusive package that is designed to wreck and destroy portable computers and laptop systems. samples of common malware include viruses, worms, Trojan viruses, spyware, adware, and ransomware. Malware may be a term accustomed describe malicious package, also as spyware, ransomware, viruses, and worms. Malware breaches a network through a vulnerability, typically once a user clicks a dangerous link or email attachment that then installs risky package. Once at intervals the system, malware can do the following:

- Installs malware or additional harmful package.
- Covertly obtains information by transmission information from the drive (spyware)

A threat in form of package is installed in our system without or permission. Various types of malwares are listed below:

1) Macro viruses: These viruses infect applications like Microsoft Word or surpass. Macro viruses attach to associate application's information sequence. once the applying is opened, the virus executes directions before transferring management to the appliance. The virus replicates itself and attaches to varied codes at intervals the computer system.

2) File infectors: File infector viruses generally attach themselves to viable code, such as .exe files. The virus is place in once the code is loaded. Another version of a file infector associates itself with a file by creating a virulent unhealthiest file with identical name, but an .exe extension. Therefore, once the file is opened, the virus code will execute.

3) Boot-record infectors: once the system is started, it will explore the boot sector and freight the

virus into memory, wherever it is regarding propagate to varied disks and computers.

4) Polymorphic viruses: These viruses conceal themselves through variable cycles of cryptography and secret writing Initially the malware is started using a secret program. Then it starts infecting the system.

5) Stealthy viruses: This infects the system using a malware detection code package. It changes the date and time of last modified state.

6) Trojans: This malware hides in each high degree programs and malicious activity is performed. The difference between virus and Trojan is that Trojans wont replicate like viruses.

7) Logic bombs: Here a package of malicious code is used. It is associated with an application and triggered simultaneously.

8) Worms: Worms disagree from viruses throughout this does not attach to numerous files. It comes through email attachments. The email attachment activates the malicious program. It replicates in each email contact address and do malicious activities.

9) Droppers: This involves an instrument containing a program that install viruses on computers. In some cases it is not infected by the virus and then it is not identified by the virus scanning program.

10) Ransomware: Ransomware are getting to be a type of malware that blocks access to the victim's info and threatens to publish or delete it unless a ransom is paid. In some cases it will lock the system fully Approaches like cryptoviral extortion is used and then it becomes hard to recover the system.

11) Adware: Adware are getting to be a code package application utilized by corporations for promoting purposes; It is mechanically downloaded to our system through popup windows or browsers.

12) Spyware: This is a type of program used to combine the data relating to user, their system and browsing history. It installs several malware programs from internet.

D. Social Engineering

Social engineering are getting to be an idea of action that adversaries use to trick you into revealing sensitive info. Social engineering is combined with any of the threats listed over to make you varied potential to click on links, transfer malware, or trust a malicious give.

E. Point of Sale Attack

Hacker uses this attack to get the customer's personal details and confidential information. The recent attack was on "Target" and forty million customers got exposed

F. Man-in-the-middle attack

Man-in-the-middle (MitM) attacks, put together cited as eavesdropping attacks, occur once attackers insert themselves into a two-party act. Once the attackers interrupt the traffic, they are planning to filter and steal info. Once malware has broken a tool, associate human will install package to technique all of the victim's information.

This threat happens when a hacker comes in between client and server. Various common attacks are:

1) Session hijacking

In this man of middle attack, a session is hijacked between a trustily shopper and network. Initially it connects with the server. Then the attacker's system get control. He disconnects the shopper from the network. Then attacker's system is replaced with client's address . But server thinks it is the client that it is communicating .

2) IP Spoofing

IP spoofing is used by associate human to influence a system that it's act with a renowned, trustworthy entity and supply the human with access to the system. The human sends a packet with the information technique supply address of a renowned, trustworthy host rather than its own science supply address to a target host.

G. Denial-of-service attack

Unlike attacks that unit of activity designed to alter the human to know or increase access, denial-of-service does not provide direct blessings for attackers. for many of them, it is enough to possess the satisfaction of service denial. However, if the attacked resource belongs to a business competition, then the profit to the human is also real enough. Another purpose of a DoS attack is to require a system offline therefore a singular quite attack is launched. Different types of DoS and DDoS attacks; SYN flood attack, teardrop attack, smurf attack, ping-of-death attack and botnets.

1) TCP SYN flood attack: This attack uses buffer through a TCP (Transmission control protocol) session acknowledgement. It responds only once the target system starts replying to requests.
2) Smurf attack: This attack involves exploitation bailiwick spoofing and place on the ICMP to saturate a target network with traffic. This attack technique uses ICMP echo requests targeted at broadcast bailiwick addresses. Another varied would be to assemble the tip systems to remain them from responding to ICMP packets from broadcast addresses.

H. SQL Injection Attack

It happens once a outlaw executes a SQL question to the info via the personal file from the patron to server. SQL commands unit of live inserted into data-plane input (for example, instead of the login or password) therefore on run predefined SQL commands. A winning SQL injection exploit can browse sensitive data from the info, modify (insert, update or delete) data, execute administration operations (such as shutdown) on the info, recover the content of a given file, and, in some cases, issue commands to the code.

The vulnerability to the present moderately cyber security attack depends on the actual undeniable fact that SQL makes no real distinction between the management and data planes. Therefore, SQL

injections work for the foremost [*fr1] if an internet machine uses dynamic SQL. to boot, SQL injection is incredibly common with PHP and ASP applications because of the prevalence of older purposeful interfaces. J2EE and ASP.NET applications unit of live less on the face of it.

The code that is dead against the info ought to be compelled to be compelled to be durable enough to forestall injection attacks. to boot, validate file against a whitelist at the appliance level.

I. Zero-day exploit

A zero-day exploit hits once a network vulnerability is alleged but before a patch or resolution is enforced. Attackers target the disclosed vulnerability throughout this window of time. Zero-day vulnerability threat detection requires constant awareness.

J. DNS Tunnelling

DNS tunnelling utilizes the DNS protocol to talk non-DNS traffic over port fifty 3. It sends machine-readable text transfer protocol and different protocol traffic over DNS. There unit varied, legitimate reasons to utilize DNS tunnelling. However, there are a unit malicious reasons to use DNS Tunnelling VPN services. they are going to be accustomed disguise outgoing traffic as DNS, concealing data that is typically shared through an online affiliation. For malicious use, DNS requests unit manipulated to exfiltrate data from a compromised system to the attacker's infrastructure. It can also be used for command and management call-backs from the attacker's infrastructure to a compromised system.

K. Drive-by attack

Drive-by transfer attacks unit a regular methodology of spreading malware. Hackers look for insecure websites and plant a malicious script into machine-readable text transfer protocol. A drive-by does not think about a user to do to one thing to actively alter the attack — you don't need to be compelled to click

a transfer button or open a malicious email attachment to become infected. A drive-by transfer can make the most of associate app, software package or browser that contains security flaws because of unsuccessful updates or lack of updates.

L. Secret Attack

Passwords unit the foremost commonly used mechanism to attest users to associate system, obtaining passwords could also be a typical and effective attack approach. Access to a person's secret could also be obtained by attempting around the person's table, "sniffing" the affiliation to the network to amass unencrypted passwords, exploitation social engineering, gaining access to a secret info, or outright plan. The last approach could also be drained either a random or systematic manner:

Brute-force secret plan implies that using a random approach by creating an endeavour fully totally different passwords and hoping that one work Some logic could also be applied by creating an endeavour password related to the person's name, job title, hobbies or similar things.

In a reference attack, a reference of common passwords is utilized to aim to realize access to a user's portable computer and network. One approach is to repeat associate encrypted file that contains the passwords, apply an analogous cryptography to a reference of commonly used passwords, and compare the results.

In order to safeguard yourself from reference or brute-force attacks, you would wish to implement associate account resistance policy which can lock the account once variety of invalid secret makes an endeavour.

M. Cross-site scripting (XSS) attack

XSS attacks use third-party internet resources to run scripts among the victim's browser or scriptable application. Specifically, the offender injects a payload with malicious JavaScript into a website's info. once the victim requests a page from the net

website, the net website transmits the page, with the attacker's payload as a vicinity of the machine-readable text nomenclature body readable text transfer protocol request before reflective it back. make sure all data is valid, filtered, or free before tinkling one thing back to the user, just like the values of question parameters throughout searches. Convert special characters like ?, & /, <, > and areas to their many machine-readable text nomenclatures .
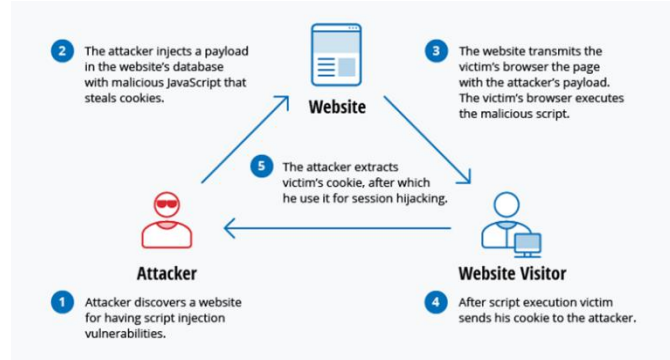
N. Eavesdropping attack

In this attack the attacker can access passwords and card information's. It is of two types. They are:

Passive: Here an attacker detects the info through analysing the message transferred through networks.

Active: Here the attacker gets the data by acting as a friend of target.

O. Birthday attack

In this type of attack instant messages created by hash are used. A message processed by a hash results in creation of another message.

V. SOLUTION TO CYBER ATTACKS

Cyber Attacks on businesses appears to be inevitable- a minimum of with the prevailing scenario within the cyber landscape. however security analysts say that to an outsized extent most of those attacks area unit evitable if corporations selected to follow the below-specified steps crafted specifically to shield their enterprises against cyber-attacks.

A. The primary and foremost resolution to forestall cyber-attacks on corporations is to own a secure and complicated hardware that area unit secret protected and saved by 2-way authentication. Also, it's higher if you are doing not overlook the effectiveness of protective your physical storage disks. as a result of if neglected, then it provides a chance to anyone and everybody to run away together with your firm's sensitive knowledge.

B. Safeguard your company's hardware- A survey conducted by IDC says that the majority of the information breaches occur once taken instrumentation reaches the hands of the hackers. for example, if one in every of your staff operating in R&D loses his/her portable computer, then the corporate that he/she is functioning for will land into serious troubles if the sensitive knowledge reaches the hands of the competitors or those having villainous intentions. Thus, it is higher to stipulate some physical security methods before any untoward incident happens. Like storing the information on the cloud that is protected by multiple security layers and inculcating accountable BYOD security policies among the workers operating for your business atmosphere.

C. Encipher knowledge- coding of knowledge provides your company associate favourable position once your data falls into wrong hands. which is thanks to the actual fact that it becomes useless notwithstanding a hacker sniffs it out- and mind you it is not that simple to interrupt into the coding on the market within the market currently.

D. Backup knowledge- typically regardless of however laborious you try; hackers get into your network and check out to encipher your data with ransomware. however, if your enterprise contains a backup copy of the most recent, then you or your company need not bow to the stress of the hackers. however, bear in mind, the backup ought to be exhausted an efficient manner which

too ought to be within the recoverable type as shortly as a disaster strike.

E. It is wise invest in cybersecurity insurance currently- as a result of cybercriminals are getting too refined these days, they are developing with ways in which to interrupt into the foremost advanced cyber defense. Therefore, even the foremost security-conscious businesses get liable to cyber-attacks wherever a cyber insurance cowl can return to your rescue. If in case, associate attack happens, most of the policies not solely cowl the loss caused from knowledge stealing however additionally facilitate in co-paying the prices concerned in ill knowledge which includes paying to knowledge recovery consultants and for purchasing new hardware still as package.

F. Educate staff on the most recent happening within the cyber landscape, so they will facilitate mitigate cyber risks with ease. This includes educating them regarding risks related to victimization unsecured networks to access work information and avoiding unsecured websites and sharing sensitive knowledge on social media. proscribing them from secret sharing will facilitate.

G. Use of anti-malware solutions and protective enterprise networks with economical firewalls will facilitate keep your enterprise IT safe from attacks.

The other solutions include:

- Train your workers.
- Keep your package and systems up thus far.
- Ensure termination Protection.
- Install a Firewall.
- Backup your knowledge.
- Control access to your systems.
- Wifi Security.
- Access Management
- Password Management
- Developing cyber security policies

- Implementing security awareness coaching
- Installing spam filters and anti-malware package
- Deploying Next-Generation Firewalls (NGFW)
- Installing termination detection & response
- Performing vulnerability assessments
- Conducting routine penetration testing
- Implementing security info and event management (SIEM)
- Deploying intrusion detection & stop package (IDS and IPS)
- Creating an information loss bar (DLP) program

VI. CONCLUSION

Computer security could be a large topic that's changing into a lot of necessary as a result of the planet is changing into extremely interconnected, with networks being employed to hold out crucial transactions. law-breaking continues to diverge down completely different ways with every New Year that passes so will the safety of the knowledge. the most recent and tumultuous technologies, at the side of the new cyber tools and threats that return to lightweight day after day, area unit difficult organizations with not solely however they secure their infrastructure, however they need new platforms and intelligence to try and do, therefore. there is no good resolution for cybercrimes, however we should always strive our maximum to attenuate them so as to own a secure and secure future in cyber house.

during this paper, the state of the art of the cyber security field of study and its importance in everyday on-line activities is investigated. It provides a helpful cyber-attack taxonomy and classification that helps users involve themselves during a protection method to spot attacks and measures for cyber security. Existing protection systems that target cyber threats and risks area unit evaluated against 3 of our criteria for an efficient anti-cyber-crime system: resilience to cyber attacks' countermeasures; real time support and needs-based action; coaching and academic materials to extend users' awareness of cybercrimes. This

analysis and review of the present attacks and tools can facilitate researchers within the cyber security field of study to propose helpful and effective protection schemes for current and future attacks.

VII. REFERENCES

[1]. Study of cyber security challenges and its emerging trends on latest technologies By G. Nikhita Reddy, G.J. Ugander Reddy

[2]. A Study of Current Scenario of Cyber Security Practices and Measures By Rajesh Mohan More, Rd. Ajay Kumar International Journal of Engineering Research and General Science Volume 2, Issue 5, August-September 2014 ISSN 2091-2730

[3]. Cyber security: Study on Attack, Threat, Vulnerability By Tushar P. Parikh, and Dr. Ashok R. Patel International Journal of Research in Modern Engineering and Emerging Technology ISSN: 2320-6586

[4]. Cyber Attacks, Countermeasures, and Protection Schemes– A State of the Art Survey By Antesar M. Shabut, K T Lwin, M A Hossain International Conference on Software, Knowledge, Information Management & Applications (SKIMA)

[5]. Analysis of Cyber Attacks and Security Intelligence By Yang Kim, II Kim, Namje Park Electrical Engineering book series (LNEE, vol 274)

[6]. A survey of emerging threats in cybersecurity By Julian Jang Jaccard, Surya Nepal Journal of Computer and System Sciences Volume 80, Issue 5, August 2014, Pages 973-993

[7]. Cybersecurity By R.A. Kemmerer IEEE Xplore: 28 May 2003, 25th International Conference on Software Engineering, 2003. Proceedings. ISBN:0-7695-1877-X

[8]. Cyber-attacks: the fog of identity V. A. Greiman IEEE Xplore: 02 February 2017 2016 International Conference on Cyber Conflict (CyCon U.S.) ISBN:978-1-5090-6172-3

[9]. From prediction to anticipation of cyber attacks Michael Weiss IJBT (2018) 01: 1-11 Research Article

[10]. Cybersecurity Research Papers Sans Technology Institute

[11]. Cybersecurity: The Hacker Proof Guide To Cybersecurity, Internet Safety, Cybercrime, & Preventing Attacks By Trust Genics Released: Jan 13, 2020 ISBN:9781393623205

[12]. Targeted Cyber Attacks: Multi-staged Attacks Driven by Exploits and Malware By Aditya Sood and Richard Enbody PUBLISHER: Elsevier Science Released: April 18, 2014 ISBN:9780128006191

[13]. Introduction to Cybersecurity By IntelligentHQ.com Released: May 25, 2016 ISBN:9783958800762

[14]. Cybersecurity: A Simple Beginner's Guide to Cybersecurity, Computer Networks and Protecting Oneself from Hacking in the Form of Phishing, Malware, Ransomware, and Social Engineering By Quinn Kiser Released: Mar 20, 2021 ISBN:9781393631217

[15]. Cyber Attacks: Protecting National Infrastructure By Edward Amoroso Released: Dec 20, 2010 ISBN:9780123849182

[16]. Cybersecurity: The Beginner's Guide: A comprehensive guide to getting started in cybersecurity By Dr. Erdal Ozkaya Released: May 27, 2019 ISBN:9781789806939

[17]. Cybersecurity – Attack and Defense Strategies - Second Edition: Counter modern threats and employ state-of-the-art tools and techniques to protect your organization against cybercriminals, 2nd Edition. By Yuri Diogenes and Erdal Ozkaya Released: Dec 31, 2019 ISBN:9781838822217

Number Plate Scanner

Vipul Sharma[1], Yatharth Mehta[1], Manushi Dhungana[1], Ms. Tinu N. S[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, India

ABSTRACT

Number Plate Scanner (NPS) helps to identify vehicle license plates in an efficient manner without the need for major human resources and has become more and more important the recent years. There are several reasons why their importance has increased. There are a growing number of cars on the roads and all of them have license plates. The rapid development in digital image processing technology has also made it possible to detect and identify license plates at a fast rate. The whole process may be done in less than 50ms. This gives 20 frames per second which is enough to process real-time video streams. Identification of vehicles is useful for many different operators. It can be used by government agencies to find cars that are involved in crime, look up if annual fees are paid or identify persons who violate the traffic rules. U.S., Japan, Germany, Italy, U.K and France are all countries that have successfully applied ALPR in their traffic management.

Keywords : Vehicle, License, Number Plate Scanner, Frames, Real Time.

I. INTRODUCTION

The system allows the users to register the license plate number of their cars either through a mobile application or through SMS along with the parking time they want to pay for. If the parking attendant wants to check if a car has a valid parking ticket, the attendant must manually enter the license plate number to look it up in a database.

The ALPR system's task is to recognize the license plate from the image or video stream, look it up in a database and see if the license plate is valid. As ALPR has been such an important task to solve the last thirty years, a number of ALPR systems already exist with varying degree of accuracy and speed.

The complexity in recognizing license plates in the different test sets will significantly impact the accuracy making direct comparisons of the accuracy without considering the complexity meaningless.

Comparing the speed performance is easier, even though there are factors impacting it, especially the pixel resolution of the license plate images.

Fig 1.1 Number Plate

II. LITERATURE SURVEY

- Anuja, P. N., 2011. License Plate Character Recognition System using Neural Network. International Journal of Computer Application. In this paper, Method for character recognition of license plate image based on two Neural Network techniques and two feature extraction approaches are proposed. It is observed that, as fan beam feature extraction method has more features for training the neural network thus its simulation accuracy is higher.

- AmrBadr et al. In this paper Automatic recognition of car license plate number got to be indispensible part in our day by day life. This paper mainly explains an Automatic Number Plate Recognition System (ANPR) using Morphological operations, Histogram manipulation and Edge discovery Techniques for plate localization and characters segmentation. Artificial Neural Networks are used for Character classification and recognition.

- Abd KadirMahamad In this paper they explained an automatic number plate inspection of letter sets of plate using image processing and optical character recognition. An imperative system has been created of training interface using LABVIEW software.

- Kuldeepak et al. In this paper, they introduced that high level of precision has been required by the number plate recognition when streets are occupied and number of vehicles are passing through. In this paper, by optimizing different parameters, they have accomplished an exactness of 98%. It is essential that for the tracking stolen vehicles and monitoring of vehicles of an exactness of 100% can't be bargained with. Therefore to accomplish better precision streamlining is required.

Fig 2-.1 Number Plate

III. PROPOSED SYSTEM

The main aim is to develop software that read the license plate content, recognize it, converts it into the text and displays in typed format. In the proposed system, the software takes the license plate in the form of pictures as an input and applies various concepts of machine learning and artificial intelligence such as deep learning, neural networks to identify and recognize the data from the user input and convert it into digitalized text format. Focuses on evaluating a person's position from the input picture. It is therefore necessary to choose the most appropriate object detection model to avoid any problems in the detection of individuals. The image of the vehicle is captured using a high-resolution photographic camera. A better choice is an Infrared (IR) camera. The camera may be rolled and pitched with respect to the license plates.

Preprocessing is the set algorithms applied on the image to enhance the quality. It is an important and common phase in any computer vision system. For the present system, preprocessing involves two processes: Resize – The image size from the camera might be large and can drive the system slow. It is to be resized to a feasible aspect ratio. Convert Color Space – Images captured using IR or photographic cameras will be either in raw format or encoded into some multimedia standards. Normally, these images will be in RGB mode, with three channels (viz. red, green and blue).

Fig 3.1 Old Scanner reading cars

The most critical process in automated license plate recognition system is the license plate extractor. In this process, we apply different techniques on image to detect and extract license plate. This process is divided in two parts.

Number Plate Detection through Haar-like features

In image processing techniques, Haar-like features[4] are used to recognize objects from image. If our proposed system is selected to detect only license plates then the Haar- like features are used for this purpose and no further processing is done. This technique is old and laborious and more over needs a large database to store the collected samples nearly about 10000 images of the plates and characters.

Number Plate Detection through Edge Detection

In the other case, if our proposed system has to recognize license plates, then the binary image is created from the image. After that following steps are performed to extract license plate from binary image:.

1. Four Connected Points are searched from binary image.
2. Width/Height ratio is matched against those connected points.
3. License Plate region is extracted from image.
4. Transformation of extracted license plate is performed.

This approach is quick and takes less execution time and memory with high a efficiency ratio.

Then the extracted license plate is passed to next component for further processing. This approach is quick and takes less execution time and memory with high a efficiency ratio.

In the further part, image processing is done on extracted license plate to remove unnecessary data. After character segmentation, the extracted license plate has only those characters that belong to license number. This also achieved with the width height ratios matching with the contours detected on extracted number plate.

Finally, the selected blobs are send to a Optical Character Recognition (OCR) Engine, which returns the ASCII of the license number

IV. SYSTEM ARCHITECTURE
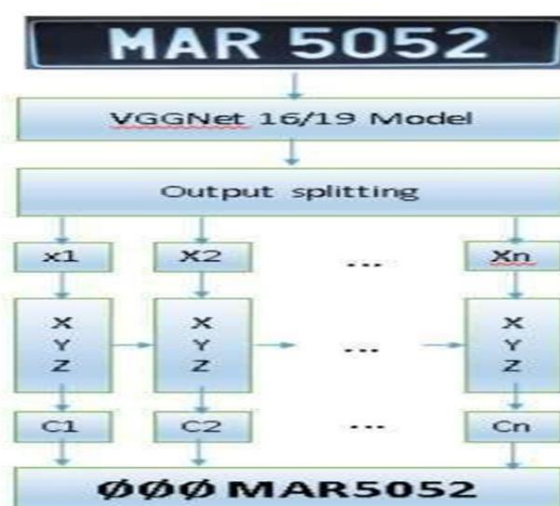
Fig 4.1 System Architecture

The system should be developed in such a format that it should be able to get the license plate in the form of picture; this picture input can be in any format irrespective of jpg, jpeg or png. After this pre-processing of the image takes place in which the entire picture is divided into small sub pictures where each picture handles individual characters and reduces the noise of the picture and converts it into its ascii values.

Next step involves in reorganization and classification of the contents on license plate with the help of various machine learning techniques such as deep

learning neural networks etc. so that accurate identification of the input is done and the last step involves in displaying the image input which will be converted in digital text format.

Fig 4.2 Sequence Diagram

V. CONCLUSION AND FUTURE WORK

Automated License Plate recognition system is mainly aiming in converting license plate data into digital text format with the help of various machine learning concepts such as deep learning, neural networks etc. and converts it into digitalized text format. the first phase of the project involved in literature survey of the project 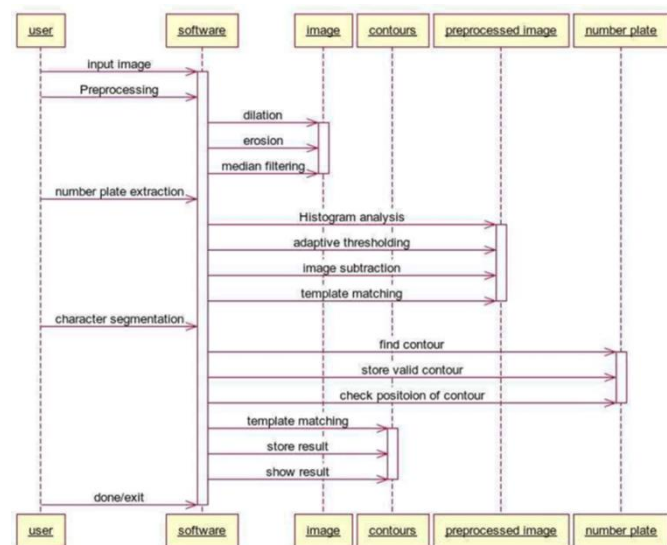identifying the methodology used to achieve the goal and it involved in giving a short presentation about the project to respective guides. Next phase involves in developing the software by writing the code in python3 with the help of APIs to meet the objective of the project and testing the efficiency and accuracy of the system.

Our approach allows the captured license plate to be passing and recognized in one forward pass and achieves high accuracy. We also prove the VGGNet in our case, deeper architecture network able to learn more discriminate feature which is robust to various illumination, rotation and distortions in the image, and lead to higher recognition accuracy.

VI. ACKNOWLEDGMENT

VII. REFERENCES

[1]. Anuja, P. N., 2011. License Plate Character Recognition System using Neural Network. International Journal of Computer Application, 25(10), pp. 36- 39.

[2]. Bharat, B., Singh, S. & Ruchi, S., 2013. License Plate Recognition System using Neural Network and Multi thresolding Technique. International Journal of Computer Applications (0975-8887), 84(5), pp. 45 - 50.

[3]. Bo, L., Bin, T. & Ye, L., 2013. Component-based license plate detection using conditional random field model. IEEE Transactions on Intellignet Transportation Systems, 14(4), pp. 1690-1699.

[4]. Guo, J. & Liu, Y., 2008. License plate localization and Character segementation with feedback self-learning and hybrid binarization techniques. IEEE Transport Vehicle Technology, 57(3), pp. 1447-1424.M.

[5]. Hui, L. & Shen, C., 2016. Reading Car License Plates Using Deep Convolutional Neural Networks and LSTMs Australia:arXiv, pp. 1-17, Technology (ICST)

Heart Disease Prediction Using Machine Learning Technique

Harshitha M[1], Dr Clara Kanmani[2]

[1]M-Tech Student Department of Computer Science and Engineering, NHCE, Maharashtra, India
[2]Assistant Professor, Department of Computer Science and Engineering, NHCE, Maharashtra, India

ABSTRACT

As the use and the advancement in technology is increasing in today's world, this leads to the exposure of new technologies to detect heart disease, which in turn leads to the increase in detection of disease as early as possible. Hence, there are many technologies to detect the heart disease but some may be time consuming and few may be costly. Hence the Machine Learning (ML) algorithm will help to detect the disease at early and with low cost. In the paper we are understanding about the machine learning algorithm that are used in detection of heart disease.

Keywords— Heart disease, Machine Learning, diagnosis.

I. INTRODUCTION

The challenging issue of the Healthcare Industry is that it faces a superiority of facility. The correct diagnosis of disease & providing effective treatment to the patients will define the way of service and its quality. The inadequate diagnosis will cause a disastrous consequence which will not be accepted. The data of medical history is very large, but these are obtained from many of the dissimilar foundations. Interpretations which are done by the physicians are the most essential components in data. The data in the real world would be inconsistent and incomplete, noisy, so the data preprocessing would be required to complete the omitted values in database. In ancient years cardiovascular diseases was found as the most source of death in world, these are been considered as the most avoidable and the manageable diseases. The accurate management of the disease will rest in the way of well-timed judgment of the disease. A proper and the methodical tool for the recognizing high-risk

patients. The mining data is required for timely analysis of the heart infection looks has been a serious want.

Even though the heart disease is considered as the supreme chronic sort of disease in the world, that can be also avoidable at the same time by proper diagnosis. A timely analysis (inferior prevention) and healthy way of life (main prevention) are the ones that should be considered as the main solution. Conducting proper check-ups (that is the inferior prevention) will shows a outstanding role in judgment and also in the early control of the heart disease. Several tests consisting of echocardiography, chest x-rays, angiography, and the exercise tolerance test that will support to the significant issue. These tests are the costly ones and that may contain the availability of the accurate medical equipment. Researchers will make use of the several data mining techniques that are available to help the physicians or specialists to identify the heart disease. The commonly used procedures used are the decision tree,

k-nearest, and Naïve Bayes. There are many other different types of the classification-based techniques used they are bagging algorithm, the kernel density, straight Kernel self-organizing map, and sequential minimal optimization and neural networks, and SVM (Support Vector Machine).

The Machine learning algorithms are the programs that will not require human intervention and learn itself from the data and help to improve itself from the experience. The Learning tasks which will contain learning the function and also will map the output from the input, learning from the concept of hidden structure in the data of unlabelled or from the Instance-Based learning, the class label would be obtained to a new instance. That will be stored by the help of comparing the value of new instance (row) to the instances from training data. The Instance-based learning will not create the abstraction from the specific instances. The Machine learning will allow doctor to speed up the process of heart detection and it will provide the accurate results and also to put preventative measures in place.

Algorithms of Machine learning are those which contain programs which does not require the help of the human intervention and hence they can learn by itself from the data and will improve from the experience. The concept of learning tasks will include the concept of learning, that will help to map the output from the input, learning structure which is hidden is unlabeled data, or the Instance-based learning, in which the class label will be used to obtain new class of instances that will be stored as comparing values of the new instance (row) to the value of instances from set of training data. The Instance-based learning will not create the abstraction for the specific of instances.

Hence, in this paper it is discussed about how the machine learning algorithms helps in detecting the heart disease will occur or not. The algorithms used are the Random Forest, Support vector machine, logistic regression, k-Nearest Neighbor (kNN), Decision tree. In this paper the accuracy of every algorithm is calculated and can get to know which algorithm provides best result and the specificity and sensitivity.

II. LITERATURE SURVEY

These algorithms will combine beliefs of three classifiers: Naïve Bayesian, Decision Tree, and k-Nearest Neighbor (kNN). Then these three algorithms showed the effect to get the value than the individual algorithm used. That is, the way in which they combine all the three specified algorithms to get the effective result [1]. ECG is the test that will measure the activity, and it will provide the valuable information regarding the heart's status. The paper tells how the classification method will be used for extracting multi-parametric features which will be obtained by analyzing HRV from ECG. Also tells about the pattern of the heart disease. The method used is the classifier which will be based on efficient FP-growth method. They conducted the experiment for the associative classifier that will utilize a set of multiple rules, and the pruning, and along with the biased of confidence [4]. These papers will present about the new heuristic for the computing efficiently sparse kernel in the SUPANOVA. This paper has applied the dataset of the benchmark Boston housing market. The Non-invasive measurement from the heart activities that will be based on magnetic field obtained by the human heart. From this prediction of the data, 83.7% of the predictions was correct. The exceeding of results obtained by using the equivalent kernels and standard Support Vector Machine. Equally in that good results was obtained by spline kernel on the benchmark Boston housing market dataset [7].

Since, early 1980's, there is a growth in Hospital Information systems (HIS) for storing the huge amount of the data. Thus, data mining methods will help for obtaining the patterns which are interesting from the databases for the reuse of data that has been stored. This is necessary in the field of the medical

research and practice, as the human beings will not be able to deal with large amounts of the data. The article will focus on characteristics of the medical data and then how the data mining will help to store the large amount of data [8]. Hand and foot disease along with the tetanus is the serious infection that occurs in less income countries. This type of infection will often affect the larger number of infants along with the young children. The death will be due to the cause of the automatic nervous system dysfunction. Early detection of disease is much difficult [10]. At present the Heart failure disease has become one of the multifaceted diseases affecting huge amount of people worldwide. At initial stage Hospitals and the cardiac centers would depend on the ECG, which is used as general tool for the evaluating and the diagnosing of disease [12].

III. MATERIALS AND METHODS

A. DATASET

A Dataset of individuals had been selected based on their history of heart problems. Based on the analysis it is found that the greatest number of deaths occur in the middle-aged people are due to the heart diseases. The dataset is select which contains 13 attributes and 300 patients records that helps in predicting the heart disease will occur or not. This data set contains records of different aged group peoples. The dataset contains attributes like age, sex, chol, restecg, thaclac, exang, oldpeak, ca,slope thal of the patient that used for identifying whether the person is at risk or not. This dataset is taken from a repository. The obtained records are split into two parts that will be Training and Testing dataset. This dataset is required for detecting the person gets heart disease or not and also to compare which algorithm works well.

B. MACHINE LEARNING ALGORITHM

1) Random Forest

It is a classifier, which is ensembled classifier using many decision trees models. It is a collection of unpruned CARSTS that follows some designed rules for tree growing. A random seed is that will be chosen from the pulls out of the random collection of samples that is in the form of the training dataset. There is a huge advantage in the random forest where it can be used in the concept of both classification and concept of regression problems. The below figure shows how random forest would look like.
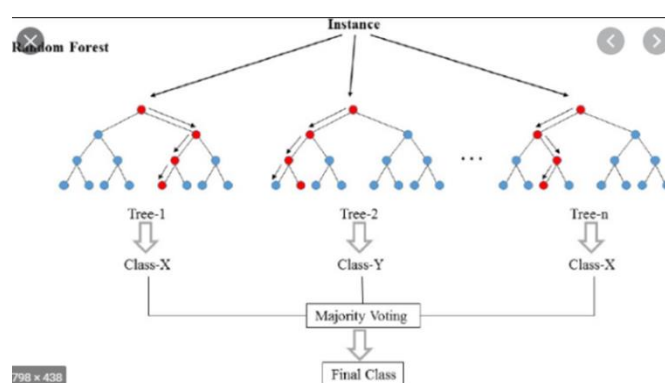
Fig.1. Random forest

2) Logistic Regression

Linear regression will help to predict values as the continuous values whereas the logistic regression will prediction will be in a discrete value after transformation function is applied.

The Logistic regression is defined for the binary classification: that is the value y = 0 or 1, 1 will represent the class that is default. For an example, in the prediction of whether a seminar will happen or not, in that there are will be only two way of possibilities, that it happens (denote as 1) or that it does not happens (0).

In the below Figure, to obtain whether the tumour will be malignant or not, then the default variable would be y is 1 then it will be tumour will be equal to malignant. The x variable that will be measure of tumour that is the size of tumour. From the figure, logistic function will be able to transform the x-value of various instances of the data set, to the limit of the value 0 to 1. The probability of value crosses the value

of threshold 0.5 (shown as the horizontal line), tumour would be classified as the malignant.

Fig.2. Logistic Regression

The logistic regression of equation is of form *P(x) = e ^ (b0 +b1x) / (1 + e (b0 + b1x))* can be transform into the form *ln(p(x) / 1-p(x)) = b0 + b1x*. The aim is logistic regression will be use the available data that are used for training and then to find the values of the coefficients that are b0 and the b1 in a way that it will help to minimize the error from the predicted value and to actual outcome.

3) Decision Trees.

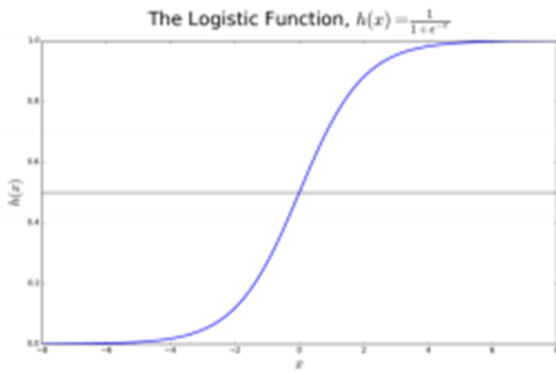The Non-terminal nodes of the Regression and Classification Trees will be root node and other will be internal node. In that, the terminal nodes are leaf node, every nonterminal node will represent the input variable that is X and also splitting point for the variable, the leaf node represent output variable, go through the entire tree to arrive the leaf node and then output the value arrived at leaf node. The decision tree shown below will classify the person whether he would buy a car or the minivan based on their age and the marital status. If person is above 30, and that person is not married, when we walk over the tree, as shown below follows that over 30 years? -> yes -> married? -> no. Hence, this will result as a sports car.
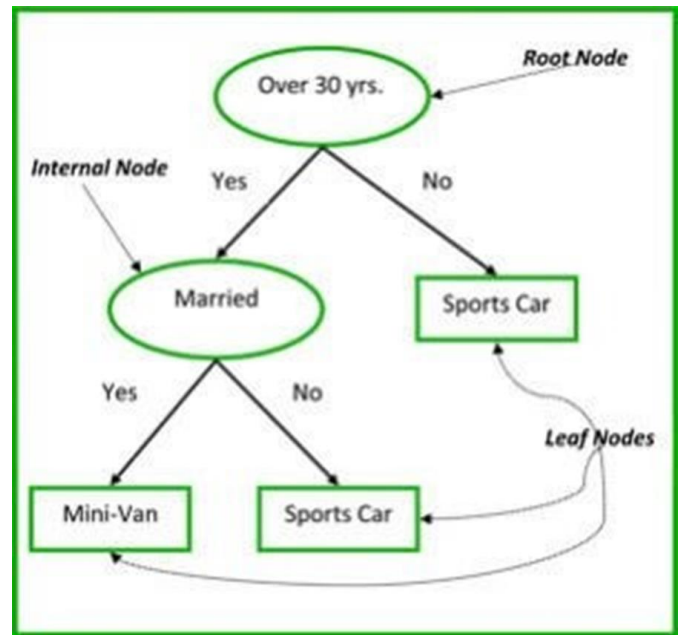
Fig.3. Some Parts of a decision tree.

4) K- Nearest Neighbour

Algorithm will use the whole set of data as the concept of be training set, instead of the splitting of data set into the concept of be training set and the other one is test set. When the outcome is needed to obtain data instance of new data, then the concept of the KNN algorithm will help to go through the set of the every data to obtain the k nearest value of instances for fresh instance or a k number of the instances will be the most identical to the new record, and then the outputs that will be mean of outcomes. The value of k will be specified by the user.
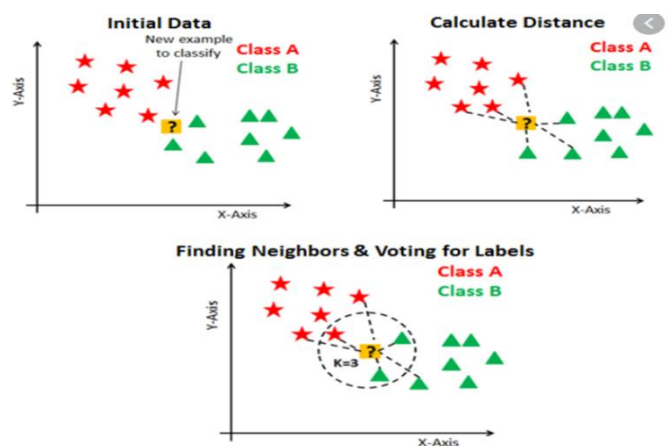
Fig.4. KNN algorithm.

5) Support vector machines (SVMs)

An SVM mode! is a concept of the different classes that is represented in concept of hyperplane in multidimensional hyperplane in multidimensional space. In the hyperplane that generated will be in an iterative manner that will be formed by SVM and so that the error that is occurred can be minimized. The goa! of the SVM will be to divide the datasets into the concept of classes and then to find a maximum marginal hyperplane (MMH). It is a type of supervised algorithm and also can be used in the concept of both regression or classification challenges. In this algorithm, first the plots of each of the data item will be as a point in the n- dimensional space (which means n represent the number of the features that contain) along with the value of every feature that will be value of coordinate that will be particular. Then, on that classification is performed and this can be done by identifying the hyper- plane. This hyper plane will help to differentiate between the two classes.
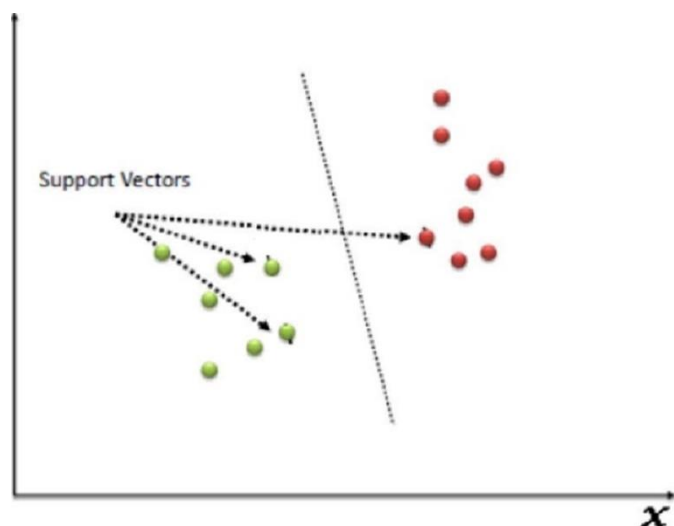
Fig.5. SVM Scenario

The SVM is one which has a 'frontier that will segregates the two classes that is hyper-plane/ line. Identify the plane that is the right hyper-plane in the following condition:
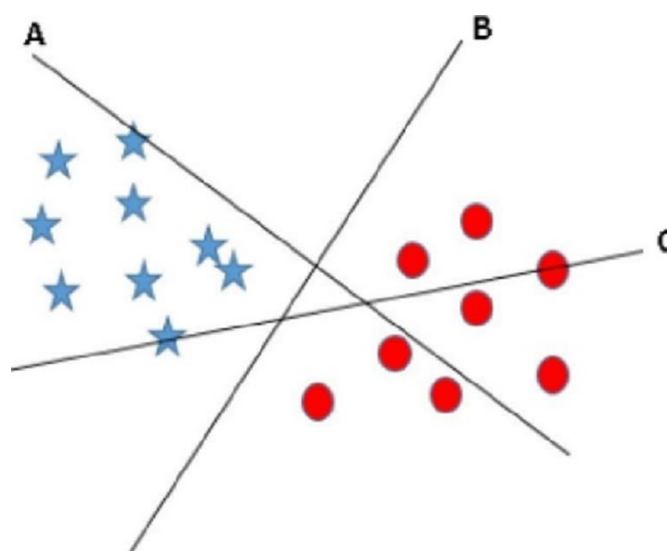

Fig.6. SVM algorithm Scenario

One must need to keep in mind about a thumb rule that will help to identify the plane that is right hyper-plane. First select the hyper-plane that will segregates the two the two classes a better. Here B has done the excellent job.

IV. METHODOLOGY

The paper main focus on prediction of whether a person will get heart attack or not by using the machine leaning algorithm. In this paper, the machine learning algorithm used are the Random Forest, Decision tree, k nearest neighbour, Support vector machine, logistic regression. The tool used here is the Jupiter lab which helps in statistical modelling, data visualization, cleaning and transformation, numerical simulation, machine learning and many more. A dataset has been collected which contain 13 attributes that are required for the analysis of prediction of the heart disease.

At the initial step of implementation, collection of dataset which contains attributes like blood pressure, age, sex thalach and other attributes which are helpful in prediction of the heart disease. Then the obtained data will be Data Validated and data cleaning will be performed that means the obtained data will be checked if there are duplicate values and

if there are any data missing values.Then if any data is in string it will be converted in to the required format. Then the data will be pre-processed that means the raw data that is obtained will be converted in to clean data set. The data pre-processing should be done before the data is feed into the machine learning model.
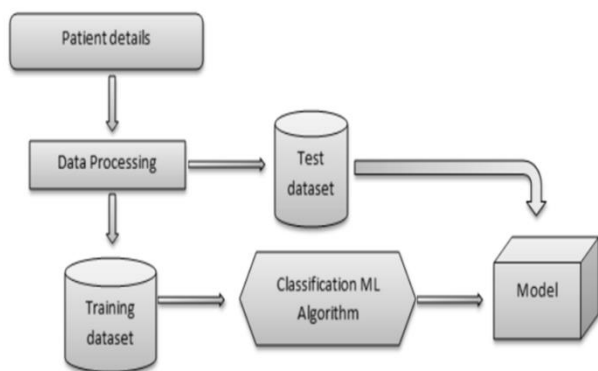
Fig.7. Architecture Diagram

In the next step we are splitting the data into training and testing data set where training set will be seventy percentage and the testing will be thirty percentage. The data visualization is also important because data sometime will not make any sense it is represented in the visual form that, means it should be in the form of charts and plots. Data visualization is important in machine learning. Then the heatmap will be plotted which helps easily to understand the differences in the values.

Next the Machine learning algorithms are used, where for each algorithm the accuracy is found individually and in this the same dataset is used to maintain the consistency that means the accuracy of each algorithm is calculated for each algorithm based on the same dataset.

Then the confusion matrix will be calculated which is used for the calculation of the sensitivity and the specificity for each of the algorithm. Then in the user interface the user will be allowed to type the attributes and predict that the heart disease will occur or not. This helps doctors also to detect at early stage whether patient will get heart disease or not.

The other parameters calculated are sensitivity and the specificity. The sensitivity value will be calculated for each algorithm that means the proportion of actual positive cases that has been predicted as positive. Specificity value will be calculated for each algorithm that means the proportion of actual negative cases that has been predicted as negative. The sensitivity and specificity values are calculated from the values of confusion matrix.

The confusion matrix contains value that are the true negative and true positive, false positive, and false negative values. True positive means correctly identified positive values who's actual and predict class is yes. True negative is correctly predicted negative value where predict and actual class are no. False positive value being actual class is yes and predict class is no. False negative value means actual class is no and predict class is yes.

The machine learning algorithms that are used are the Random Forest, Decision tree, k nearest neighbour, support vector machine, logistic regression. For each of the algorithm the accuracy is found that helps to decide the best machine learning algorithm which provides best accuracy. This model not only talks about the which algorithm will provide best accuracy but also allows the person to type the attributes and to predict that whether the person will get heart attack or not.

V. RESULTS AND DISCUSSION

Since heart disease has become most common disease nowadays it should be detected at early stage hence machine learning algorithms are used which help to detect at early stage. There are many technologies available but they are time consuming and the some may be costly. So, to overcome this machine learning algorithms discussed can be used. The machine learning algorithms that are used are Random Forest, Decision tree, k nearest neighbour, support vector machine, logistic regression.

The accuracy of all the algorithms is calculated along with specificity and sensitivity values. From the accuracy we can get to know that the logistic regression and the random forest gives the best accuracy result. The all five machine algorithms have been used to predict the weather person will get heart disease or not based on the values provided by them.

In this paper data visualization is also considered which will help to understand the some of the data values. The below figure shows how many percentages of the records are there in each type of the chest pain. In this heap matrix is also calculated to understand the percentage of records affected based on the attribute values. Here the chest pain contains five values that are 0,1,2,3,4.
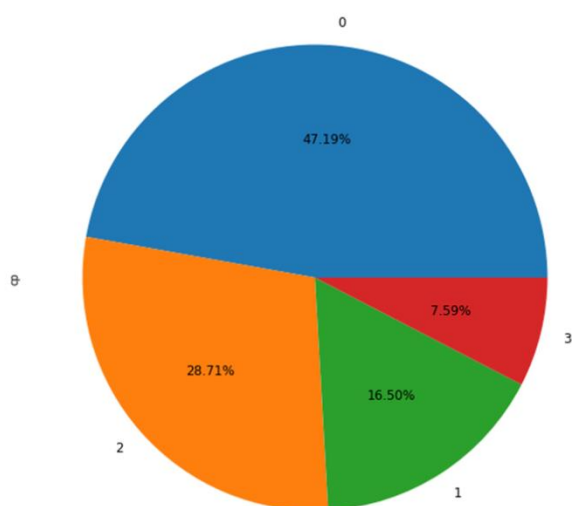

Fig.8. Percentage of chest pain

VI. CONCLUSION

This paper is presenting the detection of heart disease and the types used. Heart disease is the most common problems which occurs suddenly in the whole world. These problems cannot be seen with a naked eye and comes instantly when its limitations are reached. Hence many algorithms have been used but some are costly and time-consuming hence Machine Learning Algorithm can be used at correct time to predict the heart disease early. The analytical process will be started from the data cleaning and the processing, and the missing value, and exploratory of analysis and finally the model building and till evaluation. Finally, the prediction of the heart disease using the machine learning algorithm with different results will bring some of the following insights about heart disease prediction. As the maximum types of datasets will be covered in this will help the doctor to know about the disease exactly using ML algorithms, it helps the doctor in decision making weather patient will get heart disease or not. The machine algorithm can be used in many areas of the applications.

VII.   REFERENCES

[1]. Y. Alp Aslandogan, Gauri A. Mahajani :"Evidence Combination in Medical Data Mining".

[2]. Zhao, W., Wang, C. and Nakahira, Y. (2011) Medical Application on Internet of Things. IET International Conference on Communication Technology and Application (ICCTA 2011), Beijing, 14-16 October 2011, 660-665.

[3]. William Grant Hatcher; Weixian Liao; Weichao Gao; Wei Yu," Machine Learning for Security and Internet of Things: The Good, the Bad, and the Ugly". In IEEE access v0l 7,2019.o

[4]. Kiyong Noh,Heon Gyu Lee,Ho-Sun Shon,Bum Ju Lee,Keun Ho Ryu"Associative Classification Approach for Diagnosing Cardiovascular Disease".(volume 345)

[5]. Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Interven-tion Data" in IEEE Journal of Biomedical and Health Informatics

[6]. Soni, J., Ansari, U. and Sharma, D. (2011) Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. International Journal on Computer Science and Engineering (IJCSE), 3, 2385-2392.

[7]. Boleslaw Szymanski,arsten Sternickel,alexander Ross,lijuan Zhuu: "using Efficient Supernova Kernel for Heart Disease Diagnosis".

[8]. S. Tsumoto, "PROBLEMS WITH MINING MEDICAL DATA" Department of Medicine Informatics, School of Medicine, Shimane Medical University, Izumo, Shimane, Japan.IEEE Oct. 20000

[9]. Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal

[10]. Girmaw Abebe Tadesse; Tingting Zhu; Nhan Le Nguyen Thanh; Nguyen Thanh Hung;"Severity detection tool for patients with infectious disease"IEEE 2020

[11]. Shouman, M., Turner, T. and Stocker, R. (2012) Using Data Mining Techniques in Heart Disease Diagnosis and Treatment. Electronics, Communications, and Computers, Alexandria, 173-177

[12]. Gamal G. N. Geweld; Mahmoud A. Abdallah"A New Automatic Identification Method of Heart Failure Using Improved Support Vector Machine Based on Duality Optimization Technique"IEEE2019

[13]. Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomizd Control Trial" in JMIR Cardio 2017, vol. 1, issue 2,

[14]. Applying k-Nearest Neighbour in Diagnosing Heart Disease Pa-tients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012

[15]. Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dha-vachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[16]. Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption al-gorithm for end to end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[17]. Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B.,Vengattaraman, T., Premkumar, K., "An effective analysis on har-mony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.

[18]. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneswari Aadharshini, A., "Assessment on authentication mechanisms in dis-tributed system: A case study", (2017) Journal of Advanced Re-search in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.

[19]. Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Re-search, 10 (3), pp. 2058-2061.

[20]. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security based cloud resource virtualization", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1458-1463.

[21]. Naïve Bayes * Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering .

Evaluating the Performance of Frequent Subgraph Mining Techniques Using Apriori Based Approach and Pattern Development Approach

JagannadhaRao D B[1], S. NagendraPrabhu[2]

[1]Research Scholar, Department of Computer Science and Engineering, Shri JJTU University, Jhunjhunu, Rajasthan, India

[2]Associate Professor, Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Seeking out for inquisitively common subgraphs in chart data may be a well-studied issue in data mining. Subgraph mining strategies center on the revelation of plans in charts that show a specific organize structure that's considered inquisitively interior these data sets. The definition of which subgraphs are inquisitively and which are not is significantly subordinate on the application. Charts have gotten to be progressively critical in modeling the complicated structures. Calculation called FSP (frequent substructure pattern mining), which makes strides the state-of-the-art calculation - gSpan. In this paper the execution of FSP was assessed with FSM calculations based on both Apriori based approach and Pattern development approach.

Keywords - FSP, FSM, gSpan, Performance, Subgraph

I. INTRODUCTION

Within the most straight forward application of subgraph mining, the objective is to discover those visit subgraphs that happen more frequently than a given limit. Ready to recognize the case of numerous charts and single charts when characterizing how the events of a subgraph are tallied, which is some of the time alluded to as the bolster of the subgraph.

Some time recently beginning with the mining, input chart information must be properly encoded. After that, the primary step towards finding visit subgraphs is producing a set of candidate subgraphs. At that point, the recurrence of each subgraph in chart dataset will be checked. Usually more often than not gone before by the pruning of the look space and

expulsion of the repetitive candidates with the objective of diminishing the look space. It is ordinarily an iterative handle: bigger candidate subgraphs are produced from littler visit subgraphs. The checking step yields the event of the each subgraph that has been checked and this data is utilized to calculate the subgraph's interestingness. Visit subgraph miningis concerned with disclosure of those subgraphs from chart dataset which have visit or various events interior the given chart dataset. Inside the composing a tremendous number of visit subgraph mining calculations have been talking about; Inside the foremost straight forward application of subgraph mining, the objective is to find those visit subgraphs that happen more frequently than a given restrain. We are ready recognize the case of various

charts and single charts when characterizing how the occasions of a subgraph are tallied, which is presently and after that insinuated to as the support of the subgraph.Some time recently starting with the mining, input chart data need to be legitimately encoded. After that, the essential step towards finding visit subgraphs is creating a set of candidate subgraphs.

At that point, the repeat of each subgraph in chart dataset will be checked. Typically frequently customarily gone some time recently by the pruning of the see space and clearing of the overabundance candidates with the objective of reducing the see space. It is routinely an iterative handle: greater candidate subgraphs are made from smaller visit subgraphs. The counting step yields the occasion of the each subgraphthat has been checked and this information is utilized to calculate the subgraph's interestingness.

II. OVERVIEW OF FREQUENT SUBGRAPH MINING

This area gives a bland outline of the method of FSM. Any visit subgraph mining handle includes 3 angles , i) chart representation ii) subgraph Identification and iii) recurrence tallying.
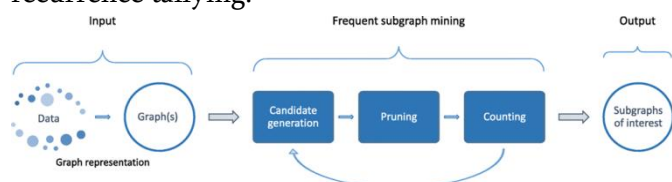

Fig1. The general procedure of finding subgraphs of internet

A. Chart representation
The only instrument whereby a chart structure can be spoken to is by utilizing an contiguousness framework or contiguousness list. Utilizing an contiguousness lattice the columns and columns speak to vertexes, and the crossing point of push i and column j speaks to a potential edge interfacing the vertexes vi and vj . The esteem held at crossing point < i, j > regularly shows the number of joins from vi to vj

B. Subgraph Identification
The current procedures for checking all the subgraphs can be classified into two categories: one is the interface operation grasped by FSG[5] and AGM [4] and another one is the development operation . The major concerns for the interface operation are that a single interface might deliver distinctive candidates which a candidate may well be unnecessarily proposed by various interface operations

C. Recurrence Tallying
Two Strategies are utilized for chart checking: Implanting records (EL) and Recomputed embeddings(RE). For charts with a single center we store an embedding list of all occasions of its title inside the database. For other charts a list is put absent of embeddings tuples that include of (1) an record of an embeddings tuple inside the embedding list of the trailblazer chart and (2) the identifier of a chart inside the database and a center in that chart. The repeat of a structure is chosen from the number of diverse charts in its embeddings list. Embedding records are quick, but they do not scale outstandingly well to colossal databases. The other approach is based on keeping up a set of energetic" charts in which occasions are more than once recomputed.

III. PROPOSED FSMALGORITHMS

The frequent subgraph discovery issue has been tended to from numerous headings utilizing different approaches, counting a priori technique and design development approach. Moreover the calculations vary within the sort of input charts, look methodology they utilize and strategy of representation of charts etc.

Consequently, there exist numerous calculations based on distinctive approaches. This makes the errand of recognizing a reasonable calculation for any given application situation an included handle. In this paper, we display a survey and propose to set up a system for classification of these calculations to assist

in understanding and analyzing different properties and restrictions of few of these calculations. A fast reference of 20 frequent subgraph mining algorithms is displayed in Table 1 and Table 2

A Frequent Substructure Pattern Mining

Han et al. have proposed many update inside the gSpan subgraph revelation calculation and named their procedure as FSP.is advanced overhaul in gSpan has finished the following success in making strides the gSpan subgraph discovery(1) moving forward the canonical chart representation utilized in gSpan and characterizing relationship between sub-DFS trees whose root shares the same parent center inside the candidate subgraphs;(2) finding the assistant resemblances inside the DFS Search Space and appearing the two one-to-one relation between children subgraph and final said kinfolk subgraphs; (3) applying the two strategies to subgraph mining issue to diminish the number of chart and subgraph isomorphism tests suitably; and (4) executing the FSP calculation and other current related methods and comparing the results.

IV. OUTCOMES

Within the taking after areas we have assessed the performanceof the FSM strategies inspected inside the think approximately. Tests were carried out on the datasets, whose real information is given in table Inside the taking after fragments we have surveyed the performanceof the FSM methods talked around inside the consider. Tests were carried out on the datasets.

TABLE I

Classification of FSM calculations based on Design development approach

S No	Algorithm	Input type	Graph representation	Subgraph generation	Frequency counting	Nature of output	Limitations
1.	SUBDUE	Single large graph	Adjacency matrix	Level-wise search	Minimum description code length	Complete set of frequent subgraphs	Extremely small no. of patterns
2.	GSpan	Set of graphs	Adjacency list	Rightmost extension	Depth first search (DFS) lexicographic order	frequent graphs	Not scalable
3.	Close Graph	Set of graphs	Adjacency list	Rightmost extension	DFS	lexicographic order Closed Connected frequent graphs	Failure detection takes lot of time over head
4.	Gaston	Set of graphs	Hash table	Extension	Embedding lists	Maximal frequent sugraphs	Interesting patterns may be lost.

TABLE 2

Classification of FSM calculations based on Apriori based approach

SNo	Algorithm	Input type	Graph representation	Candidate generation	Frequency counting	Nature of output	Limitations
1.	FSG	Set of graphs	Adjacency list	One edge extension	Transaction identifier	(TID) lists Frequent connected subgraphs	Np-complete
2.	FFSM	Set of graphs	Adjacency matrix	Merging and extension	Sub-optimal canonical adjacency matrix tree	Frequent subgraphs	Np-complete
3.	ISG	Set of graphs	Edge triplet	Edge triplet extension	TID lists	Maximal Frequent subgraphs	Incomplete set of Graphs
4.	SPIN	Set of graphs	Adjacency matrix	Join Operation	Canonical Spanning Tree	Maximal frequent subgraphs	Non maximal graphs can also be

All our tests have been performed on a 32-bitLinux system with 4GB memory and 3.0 GHz Intel processor.Executable of gSpan, Gaston has been gotten from therespective makers and remaining strategies have beenre-implemented. The execution of the inspected frequentsubgraph disclosure calculations has been surveyed on thethree datasets but gSpan whose execution is evaluatedwith respect to other procedures on one dataset, that's , compound dataset, as for the dataset where thenumber of chart increases 400, the gSpan took much time.For this reason we have compared all other strategies on three datasets but gSpan.
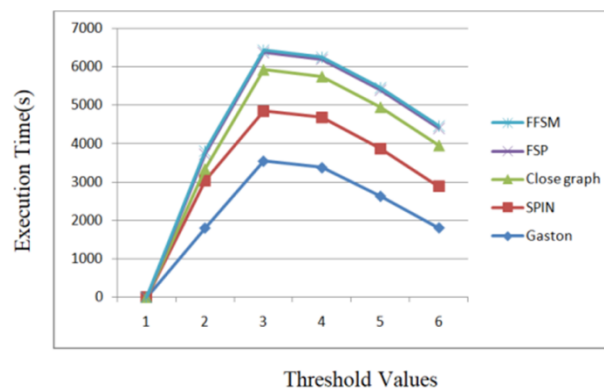

Fig 1.Comparison of various FSM techniques on DTP

V. CONCLUSION

Chart mining might be a well-explored zone of explore inside the datamining community in which the visit subgraph discoveryis one of the first challenging issues. In composing, visit subgraph

disclosure calculations havebeen proposed. this work has given a comprehensive portrayal of a number of outstandingly discernible and efficient subgraph disclosure calculations in conjunction with their essentialimplementationalgorithms.Inthisstudy,weha veprovidedacomprehensivecomparison of diverse FSM strategies and re-implemented these strategies.

VI. REFERENCES

[1]. R. Vijayalakshmi, R. Nadarajan, J. F. Roddick, M. Jilaga, andP. Nirmala, "FP-graphminer - A fast frequent pattern miningalgorithm for network graphs," Journal of Graph Algorithms andApplications,vol 15, no 6,pp 753-776, 2011.

[2]. S.Han,K.N.Wee,andY.Yu,"FSP:frequentsubstructu repat-tern mining," in Proceedings of the 9th International Conferenceon Information and Communication Security (ICICS '07), IEEE,Zhengzhou, China, December 2007.

[3]. K. Lakshmi and T. Meyyappan, "A comparative study offrequentsubgraph mining algorithms," International Journal ofInformation Technology Convergence and Service,vol.2,no.2,p.23, 2012

[4]. K. Lakshmi and T. Meyyappan, "Frequent subgraphminingalgorithms—a survey and framework for classification," inProceedings of the Conference on Innovations in ⸭eoreticalComputer Science (ITCS '12), pp. 189–202, 2012

[5]. SN Prabhu, Improving the performance of IDS using Arbitrary Decision Tree in Network Security, International Journal of Advanced Science and Technology Vol. 29 no.3, PP 3453 – 3462, 2020.

[6]. SaifurRehman&SohailAsghar, Performance Evaluation of Frequent Subgraph Discovery Techniques, Mathematical Problems in Engineering 2014:Vol-1, no -6.

Automated Tollbooth Controller Based On RFID

M Praveen Raju[1], M M Nawaz Baig[1], K Sreeram Raju[1]

[1]Department of Computer Science, New Horizon College of Engineering, Kadubeesanahalli, Marathahalli, Bangalore, Karnataka, India

ABSTRACT

This will automate the tollgate based on web application that allows user to pay digitally and the toll controllers to operate the booths more efficiently. Scanner detects the taq when it is in 30-40 meter away from the tollbooth & detects the money from the bank account and open the tollbooth gate This system includes RFID card which will be used as the access card for the users at the tollgates, making the work much more easier and with less man power.

The main motive of our group behind this project is to make an automated system for collecting the toll so that the system works faster and in a more systematic way. As per we all know that presently there are people sitting on the toll stations for collecting the toll amount and then they manually feed the data into the system and then the token is generated but by using our system this whole process can be done automated. After going through various techniques available in the present world we came to a conclusion of using RFID (Radio Frequency Identification).

I. INTRODUCTION

As we all know that the population of our country is increasing day by day the same follows with the number of vehicles on the roads. The increase in number of vehicles has not only increased the traffic on the roads but also increased the problem of the government. One of the major problem is availability of good roads for safe movement of the vehicles and for the same the government is making so many new highways and subways so that the public does not faces any issues. Now as we know nothing comes for free so the people have to pay for the same for availing these facilities. This is done by collecting tolls on the toll gates/toll stations.

The main motive of our group behind this project is to make an automated system for collecting the toll so that the system works faster and in a more systematic way. As per we all know that presently there are people sitting on the toll stations for collecting the toll amount and then they manually feed the data into the system and them the token is generated but by using our system this whole process can be automated thus reducing the expense of the government and also in faster execution of the process. After going through various techniques available in the present world we came to a conclusion of using RFID (Radio Frequency Identification). RFID tags are the tags which have been gaining the attention of people in recent days because of its speed of data handling process and processing it in the desired way and also working with the data using the methodologies such as bar coding.

In this world full of technologies, where the machines are making the way of living easier day by day we hope that our project will add up to the same. People today travel all around the world using their Own source of transport such as personal cars due to which the traffic controlling system on the highways needs an upgradation as soon as possible so that we can avoid the usual traffic jams and the long queues at the toll stations so that the traveler need not spend a lot of time and the manual collection of cash can also be automated for a better handling of the accounts. Therefore our project's main motive is to reduce the time consumption, human efforts and immediate account settlements for a faster execution of the process. This will automate the tollgate based on web application that allows user to pay digitally and the toll controllers to operate the booths more efficiently. This system includes RFID card which will be used as the access card for the users at the tollgates, making the work much more easier and with less manpower.

Smart card has not yet been penetrated in the toll booths especially in India. If this system gets inducted in toll booths of India then it will save the time and secure the transactions. As the number of vehicles are increasing due to which it makes an employee's very hectic and time consuming for the customers, this project will reduce the men power and more automated. At present people are more digitalized and going cashless it can be operated over phone, laptops tablets.

II. LITERATURE REVIEW

Charlie Fine et. al: The major driver for its development has been the tagging of physical objects people, places, and things with single chip radios so they can interface with computers.. An RFID system can be broken down into two key dimensions. The technical infrastructure includes the actual data capture technology comprised of tags, readers, and transmission medium. The logical infrastructure refers to the overall identification (ID) scheme used in representing objects.

Ahuja Sanjay and Potti Pavan: RFID is still in a developing phase and more is in the pipeline in terms of new applications. Among applications already developed, RFID tags are being used in clothing for billing and security purposes. RFID tags are embedded inside animals for tracking purposes. RFID tags embedded in uniforms can be used to know the number of hours an employee spends to complete a par-ticular task.

Matija Bumbak: RFID are surrounded with three primary issues and the need to protect proprietary information Protecting data stored on the tag; Protecting the integrity of the tag (and thus thee product);Securing data related to the serial number on a tag, which may be stored in a network database.

Kamaran Ahsan et. al: Industries use RFID for various applications such as personal/vehicle access control, departmental store security, equipment tracking, baggage, fast food establishments, logistics, etc.. Other applications includes automatic toll payments, departmental access control in large buildings. personal and vehicle control in a particular area, security of items which shouldn't leave the area, equipment tracking in engineering firms, hospital filing systems

Srivastav Nandita: RFID is an automatic identification system. RFID uses Radio frequency to identify tagged items .This data is then collected and transmitted to a host system using an RF Reader. The data transmitted by the tag may provide identification or location information, or specifics about the product tagged, such as price, color, date of purchase; etc.In Bar code the scanner device directs a light beam at the bar code. The device contains a small sensory reading element. This sensor detects the light being reflected back from the bar code, and converts light energy into electrical energy. The result is an electrical signal that can be converted into data.

III. EXISTING SYSTEM

The current system based on Active System of the active tag vehicle monitoring system of products have a range of 30-40 meters and operate in the 916-930 MHz for to recieve link. Active wave products are currently equipped with 256 kbits of memory which is also equipped with battery of power 3V battery and which totally make the weight 14 gms. Basic signals are shown with the aid of power on and power off of LEDs and beeping sounds. Smart cards access which will control systems have a client-server model based with an SQL server which will handle the database which will monitor systems. They have designed a GUI(user interface) using the Microsoft .NET framework. Smart access also operates on the 900 MHz band but have small range of 30-40 meters. RFID based toll booth collection system uses active.

RFID tag which uses solar or car battery power. The development is segregation into the design of two modules. The 2 modules(active tag) and base modem connected to each module. These RF modules communicate over the ISM frequency which ranges of 900-930 MHz.
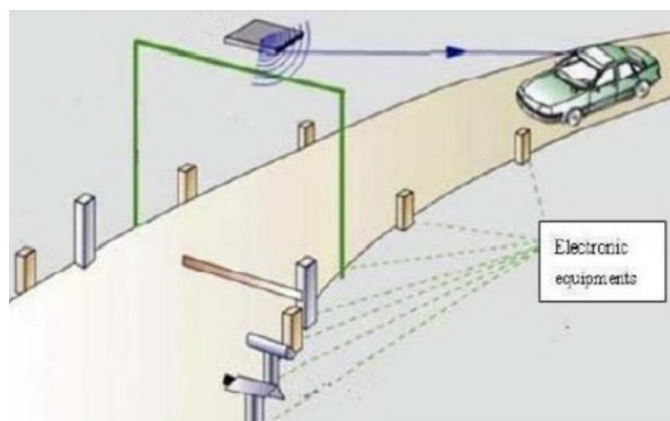

Fig :3.1

IV. PROPOSED SYSTEM

The goal of this project is the simplification of the traditional rules followed by the people or the passengers who transit the money at the toll booths, the main motto is to make the whole system to make

it automated, theft detection etc, all kinds of activities can be carried out using this single RFID card, thus saving the effort made by the passengers of carrying the money and later vehicle records manually.

This project comes with the concept that much of the modern people are looking forward to it tend to save their time in much effective way and easy to use, this automatized system will help people in saving their leisure time and more technologically advanced in every way, as we experience a lot of time consumption at the toll booths where many vehicles and spend lot of time paying and being in queue which wastes a lot of time of the modern people many concepts where proposed to ensure that no time is wasted there hence this project is a modern approach to that problem.

It also helps to efficiently cut down the pollution emission and fuel consumption and which will increase the economy of a country to a greater level this have a very advantageous effect on the social and economic life of a country, with these kind of advanced technology we can gain lot of economic advantages from it, as we know that the road ways are the major route of connectivity from ancient to a modern country so the top priority of the country should be to maintain and improve the infrastructure of the road transport.

Fig :4.1

while the modern RFID based transaction consumes only 3 seconds on an average which efficiently cut

downs the time consumption to larger fraction, and also they recorded that the normal toll booth moves only 400-550 number of vehicles every hour which is totally inefficient as compared to that of the RFID based toll booth which moves 600-800 vehicles every hour which is far more efficient than the traditional way of collecting the toll. which also helps to efficiently cut down the pollution emission and fuel consumption and which will increase the economy of a country to a greater level

V. PROJECT FEATURES

This project is consists of an android app which will manage all sort of customers and registering them to the pool, hardware is developed on RFID based and embedded C programming, which will provide a real time experience while experiencing the automized system. The android app will use the real time processing of the data accessed from the customer and will produce a desired output.

VI. LIMITATIONS

- Scanner cannot detect the tag over a certain speed.
- Regular maintenance will be required.
- Users will have to maintain enough balance at all times.
- Tag placement will have to be in the right place.

VII. FUTURE ENHANCEMENT

- We can start using this in new toll booths and slowly increase in number.
- We can eliminate the need of humans at booths.
- We can allow users to deposit money from multiple sources.

VIII. CONCLUSION

After doing a detailed study of this topic we can come to a conclusion that this system can be very beneficial not only to the government but also to the general society and the people. The major parts which are being effected by using this system can be briefed under the following benefits such as fuel saving, traffic reduction as well as saving of the environment. It will be very helpful in keeping the revenue generated by the toll safe as well as make the money reach the government without any corruption which will help in generating more revenue. This system will be very helpful for the traveler who travel on daily basis through the toll gates because they need not spend a lot of time on tolls and could also keep a track of their travel through different tolls which will lead to a proper accounting of their money too. The passenger need not pay any extra money to the people at the toll gates and have to pay the amount which has been fixed by the government. The notifications about the deduction of the money from account and the amount of money deducted will give the user a clear idea of the amount which has been deducted for using the paid highways. This system will also help in reducing the unethical movement of vehicles because all the vehicles will be having certain unique identification and if there is any unregistered vehicle happening to pass by then it can be easily notified and the actions can be taken accordingly. In our project we are storing this data in memory storage but if this system is embedded on a large scale then this can be done using database system and by adding up image processing methods then this system can prove as one of the best systems for the national security.

IX. ACKNOWLEGMENT

New Horizon College Of Engineering, Bangalore under Visvesvaraya Technological University, Karnataka, India.

X. REFERENCES

[1]. "Electronic system usingRFID tagging" byDr.Wadad Ismail

[2]. "Smart highway electronictoll collection"byGanesh K Andurkar

[3]. "A model for optimizing Electronictoll collection system"by DavidLevinson.

[4]. www.wikipedia.com

[5]. "Cost-benefit analysis of electronictoll collection" by Mohmoud

[6]. Apoorva Phaniraj, Manasa Kashyap "Arduino Based Electronic Toll Collection" published in International Journal of Innovative Research in Science.

[7]. SoniRani,"Wi-Fi Approach for Toll Tax Application"http://dspace.thapar.edu:8080/dspace/ bi tstream/1234 56789/260/1/91889.PDF

[8]. shodh.inflibnet.ac.in:8080/jspui/bitstream/12345 6 789/3779/3/03_litreature%20review.pdf

[9]. Khali, C.W. Michael, H. Shahriyar "Toll Collection Technology and Best Practices", Project 0-5217: Vehicle/License Plate Identification for Toll Collection Application, January 2007.

Design of Python Based Sales Analysis Website for Small Scale Retailers

Mohan Sai Krishna[1], Mohan Yadav[1], Uma. N[1]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Outer Ring Rd., near Marathalli, Kadubessanahalli, Bengaluru– 560103, Karnataka, India

ABSTRACT

This paper presents a framework capable of accurate analysis of real time sales data to forecast future sales, visualize sales, and draw important insights or patterns associated with products to achieve greater profits. The proposed framework that would be of great use for any company operating in the retail industry is based on Facebook's Prophet Algorithm for Sales forecast, Plotly Express visualizing tool for Sales visualizations and Fp-growth algorithm for generating association rules between products. Django web framework is used to integrate the algorithms into a user interactive website.

Keywords — Sales forecasting, Real-world dataset, Prophet, Plotly Express, Data Visualization, Fp-growth, Association rules, Django

I. INTRODUCTION

Countries with higher degree of population tend to have huge market competitiveness and it becomes difficult day by day for retailers to cope with this competition, especially of small scale retailers. Gaining attentions of customers plays a crucial part in sales which has become difficult, due to variety of purchase choices provided by the market. More so ever it has become important for the retailer to have the best first impression with the customers due to the competitiveness which could only be achieved by showing greater hospitality through emotions as well as products. Hospitality through emotion is the job of retailer but Hospitality through product is the one which is being focused in the paper.[5] This is could achieved through various means like better inventory control, better production planning etc., but "how to do it" is the question which is being answered in the paper. The paper proposes a website developed using

Django which would help retailers to visualize sales, forecast sales, and generate association rules between products that would help retailers in better production planning, better inventory planning and better store layout management. [3]The website proposed in the paper produces a product-level sales analysis report as it plays a crucial part for the retailer to introspect the aspects of the sales report and make responsible changes in store. [2]While accurate and reliable analysis reports can lead to huge savings and cost reductions by facilitating better production and inventory planning, competitive pricing and timely promotion planning, [6] poor sales analysis are proven to be costly in this domain since it is well-known that goods shortages, irregular product layout and placement cause lower profits and can easily lead to customer dissatisfaction. The report consists of three major parts [4] data visualization achieved using plotly express, then for sales [1] forecasting achieved using prophet and last but not the [7] least

association rules which are generated using fp-growth algorithm.

A. Problem Definition

The paper discusses about a framework that would help retailers to make more effective decisions in terms of inventory control, product production, shop layout etc. The framework provides three basic functionalities which includes Sales forecast, Sales Visualizations and product association rules. Sales forecast is achieved using Facebook's Prophet Package which produces almost accurate results using which retailer can foresee the sales in future thus, helping him to plan his product production and inventory control. The output produced by the package also includes Weekly, Monthly, and Yearly Sales Trend which gives Prophet Algorithm, an upper edge on Other Algorithms used in the market .Sales Visualizations is produced using Plotly Express data visualization tool. This particular tool produces Interactive Graphs of Various kind which helps the retailer the draw insights on Sales Such as Sales trends of particular product, Highest Grossing Product etc. The quality of the output produced in the above functionality, is directly proportional to the quality and quantity of the dataset used. The Last functionality which the framework provides is Product association rule generation which uses FP – growth algorithm. FP-growth algorithms is much more effective when compared to other algorithms which is mainly Apriori Algorithm, but it has its own cons, but as a Whole it is effective. The output produced by the algorithm includes a number associated with a product to another product or product to group of products or group of products to other group of products. The number explains the probability of one entity being bought when it is already stated that other entity associated with that entity is already bought. In Simple terms, if the number is associated between bread and jam, the probability of Jam being bought when he customer

has already bought bread .This functionality helps retailer to plan store layout, improve the sales of off-season products etc.

The framework proposed by the paper helps retailer to gain an extra step when compared with the other market competition, as the retailer would be better prepared and thus provides a satisfactory experience for the customer.

B. Objectives

1) The basic objective is to implement each of the three functionalities as individual units without any bugs in the units.
2) First objective is to have effective registration and signup platform for the retailers.
3) First unit to develop is sales forecast which requires very strict dataset to be uploaded and then forecast the sales using prophet package.
4) Retailers have to use forecasting technology for sales predictions and create a competitive edge over the rest of the market.
5) Second unit to develop is Sales visualization which requires very strict dataset to be uploaded and then outputs are produced using Plotly Express tool.
6) Retailers can analysis the Sales output to recognize insights in the sales like highest grossing product etc.
7) Third unit to develop is Product association rule generator which requires very strict dataset to be uploaded and then generate the rules.
8) Retailers can use the above mentioned framework to generate relationships between products and improve sales.
9) All the above discussed units are integrated in a dash application and the launched as django application.

C. Scope Of the Project

The Scope of this Project is the Quality of the outputs produced and the algorithms used to produce them. Even though the algorithms used in the project are the latest but it still has its flaws. The Facebook Prophet Package still requires a very strict dataset format. It requires a univarite time series dataset with date column titled as 'ds' and data titled as 'y' without which the package can't work. The algorithm can be extended in such a manner in which such strict dataset is not required. The quality of the output is directly proportional to quality and quantity of the dataset uploaded which could be changed in near future.

FP-growth algorithm used for Rule generation can be made to be less expensive in terms of time and space complexity.

Integrating the ML models into web app was complex due to involvement of User Interaction with the output produced was a must. In this particular Project, Django web framework and dash framework is used to produce the desired output. The complexity of integration can be reduced in the future. The framework can also be reduced to a mobile app for better accessibility for the User. The framework can be integrated into the billing systems of the Retailers which is Ideal Platform for the framework.

II. LITERATURE SURVEY

A. Technology

1) FBprophet: Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It gives the best output with time series that have strong seasonal effects and several seasons of historical data. Prophet is vigorous to missing data and shifts in the trend, and typically handles outliers well. Prophet is open source software developed by Facebook. It is available for download on CRAN and PyPI. Prophet is used in

many platforms across Facebook for producing quality forecasts for planning and goal setting. It is found it to perform better than any other approach in the majority of cases. Prophet is vigorous to outliers, missing data, and dramatic changes in your time series. The Prophet procedure includes many possibilities for users to tweak and adjust forecasts. The human-interpretable parameters can be used to improve your forecast by adding domain knowledge. Prophet has been implemented in R and Python, but they share the same underlying Stan code for fitting.
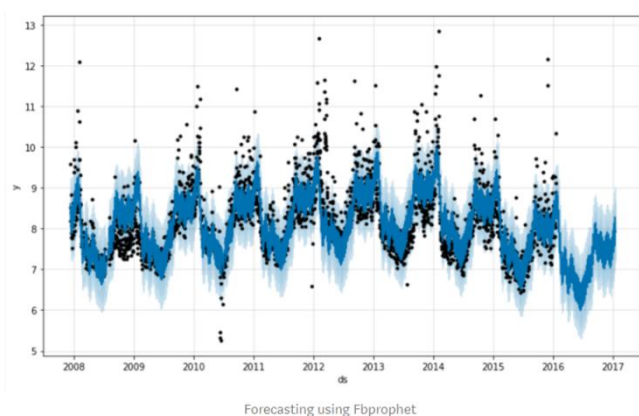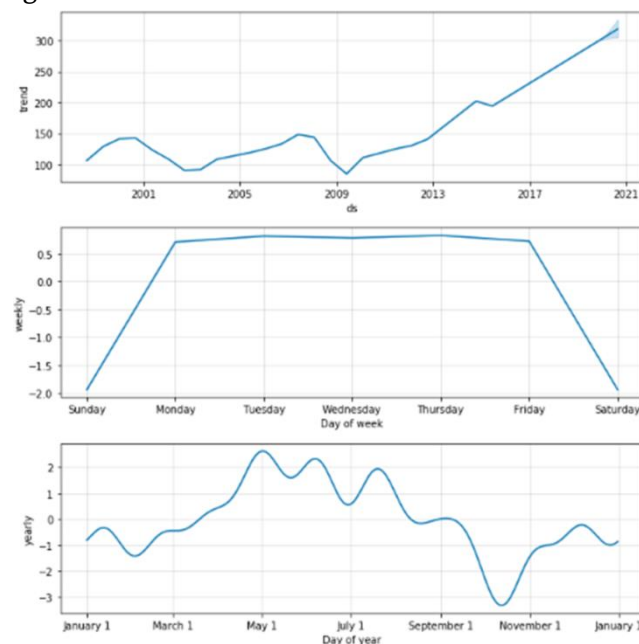

Fig.1 Sales forecast

Fig.2 Forecast Components

2) Plotly – Dash: Dash is a productive Python framework for building web analytic applications. Written on top of Flask, Plotly.js, and React.js, Dash is ideal for building data visualization apps with high user interfaces in pure Python. It is perfect for anyone who works with data in Python. Through a couple of simple patterns, Dash abstracts away all of the technologies and protocols that are required to build an interactive web-based application. Dash is simple enough that you can bind a user interface around your Python code in an afternoon. Dash apps are rendered in the web browser. You can deploy your apps to servers and then share them through URLs. Since Dash apps are viewed in the web browser, Dash is inherently cross-platform and mobile ready.

Fig.3 Dashboard developed using Dash Plotly

3) Django: Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hard work of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a growing and active community, great documentation and etc. Django follows the "Batteries included" philosophy and provides almost everything developers might want to develop an unique idea. Because everything you need is part of the one "product", it all works effortlessly together, follows consistent design

principles, and has up-to-date documentation. Django has been used to develop all kinds of websites like content driven, management etc It can work with any client-side platform, and can deliver content in almost any type of format (including HTML, RSS feeds, JSON, XML, etc). Internally, while it provides choices for almost any service you might want (e.g. several popular databases, templating engines, etc.), it can also be extended to use other components if needed. Django helps developers avoid many common security mistakes by providing a framework that has been engineered to protect the website automatically. For example, Django provides a secure way to manage user accounts and passwords, avoiding common mistakes like putting session information in cookies where it is vulnerable or directly storing passwords rather than a password hash. A password hash is a fixed-length value created by sending the password through a cryptographic hash function. Django can check if an entered password is correct by running it through the hash function and comparing the output to the stored hash value. However due to the "one-way" nature of the function, even if a stored hash value is compromised it is hard for an attacker to work out the original password. Django enables protection against many exposed by default, including SQL injection, cross-site scripting, cross-site request forgery and click jacking. Django uses a component-based "shared-nothing" architecture (each part of the architecture is independent of the others, and can hence be replaced or changed if needed). Having a clear separation between the different parts means that it can scale for increased traffic by adding hardware at any level: caching servers, database servers, or application servers. Some of the busiest sites have successfully scaled Django to meet their demands. Django code is written using principles and patterns that encourage the

creation of sustainable and reusable code. In particular, it makes use of the DRY principle so there is no unnecessary duplication, reducing the amount of code. Django also promotes the grouping of related services into reusable "applications" and, at a lower level, groups related code into modules (along the lines of the Model View Controller (MVC) pattern).Django is written in Python, which runs on many platforms. That means that you are not tied to any particular server platform, and can run your applications on many platforms which makes it portable. Furthermore, Django is well-supported by many web hosting providers.
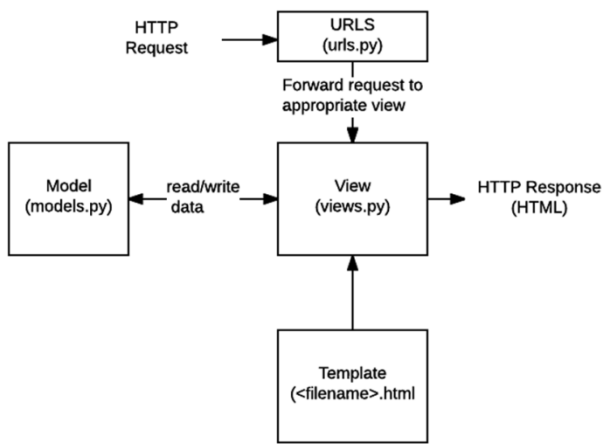

Fig.4 Django App Framework

4) FP–growth: This algorithm is an improvement to the Apriori method. A frequent pattern is generated without the need for candidate generation. FP growth algorithm represents the database in the form of a tree called a frequent pattern tree or FP tree. This tree structure will maintain the association between the itemsets. The database is fragmented using one frequent item. This fragmented part is called "pattern fragment". The itemsets of these fragmented patterns are analysed. Thus with this method, the search for frequent itemsets is reduced comparatively.
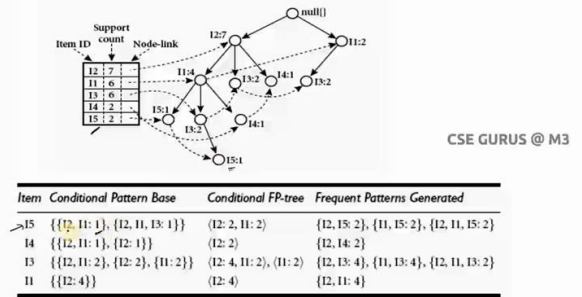

Fig.5 FP-Growth Algorithm

B. Existing System

The Existing Systems use Arima model for forecasting sales which has various disadvantages:

1) Quantity of dataset required is more when compared to prophet algorithm to quality output.
2) Unlike simple naive models or smoothing models there is no automatic updating feature as new data become available the entire modelling procedure must be repeated especially the diagnostic checking stage as the model may have broken down. The fact that these models tend to be unstable cause many economists to view them with suspicion.
3) Because of the large data requirements, the lack of convenient updating procedures, and the fact that they must be estimated using nonlinear estimation procedures, the B-J models tend to be high cost. The amount of subjective input at the identification stage also make them somewhat more of an art than a science.
4) The ARIMA model tend to be unstable, both with respect to changes in observations and changes in model specification. many specification will yield no results, and as with most nonlinear estimation techniques, the results may not be unique

User Interaction in the output was very limited in the existing system due to the absence of Ploty Express

data visualization tool. Prerequisite knowledge was required in the existing system for analysis the output report. Some amount of knowledge was also required to operate the Existing system. Integrating such ML models was difficult and much more complex in existing system when compared to the system which are developed now.

C. Methodology to be followed

To finish this project and make a working model that showcases our idea within the given deadline, we have decided to follow various Software development and Software Testing principles and methodologies:

1) Planning: The purpose of the first phase is to find out the scope of the problem and determine solutions. Resources, costs, time, benefits and other items are generally considered here. In our project we have decided to follow a divide and conquer approach to finish certain tasks, with each of us having various skill sets, we were able to almost perfectly give roles that we are most proficient at, and this planning has helped us to minimize time, and maximize utilization of resources.

2) Research Paper Analysis and Requirements: In this phase we have gone through research papers put out by other people, related to forecasting and market basket analysis, expanding our horizon on these topics, and checking the feasibility as well as limitations of our project.

3) Software Engineering Methodologies: There are various SWE methodologies, and we had to pick the one that works the best for the development of our project.

- Waterfall: The waterfall model is a linear project management approach, where stakeholder and customer requirements are gathered at the beginning of the project, and then a sequential

project plan is created to accommodate those requirements.

- Agile: The Agile methodology was developed as a response to growing frustrations with Waterfall and other highly structured, inflexible methodologies. This approach is designed to accommodate change and the need to produce software faster.

- Feature - driven development: An Agile methodology for developing software, Feature-Driven Development (FDD) is customer-centric, iterative, and incremental with the goal of delivering tangible software results often and efficiently. FDD in Agile encourages status reporting at all levels, which helps to track progress and results.

- Scrum: Another way to implement the agile approach, Scrum borrows from Agile's foundational beliefs and philosophy that teams and developers should collaborate heavily and daily. With Scrum, software is developed using an iterative approach in which the team is front and center—experienced and disciplined workers on smaller teams might find the most success with this method, as it requires self- development. As with most Agile approaches, XP allows for frequent releases in short development sprints that encourage change when needed.

- Lean: Lean is at once a workflow methodology and a mindset, incorporating principles and practices from the manufacturing space and applying them broadly to a variety of industries,

- Including software development. While Agile is an excellent methodology for the practical application of development best practices, it does not include instructions for scaling these practices across the organization or applying them outside of development- type work.

4) Software testing methodologies: Software testing methodologies are the various strategies or approaches used to test an application to ensure it

behaves and looks as expected. These encompass everything from front to back-end testing, including unit and system testing.

5) Data Collection: Data collection is the process of gathering and calculating information on varying interest, in a very established and systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluating the outcomes.

III. REQUIREMENT ANALYSIS

A. Functional Requirements

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behaviour, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioural requirements describing all the cases where the system uses the functional requirements are captured in use cases.

Here, the system has to perform the following tasks:
- User Friendly Framework that doesn't require Perquisite Knowledge.
- Produce accurate and quality forecast of sales
- To Produce User interactive output for sales data visualization
- To produce quality product association rules so, that retailer can make effective decisions.
- To integrate models into a single web app using Django and Dash.

B. Non-Functional Requirements

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. This should be contrasted with functional requirements that define specific behaviour or functions. The plan for implementing functional requirements is detailed in the system design.

1) Accessibility: Accessibility is a general term used to describe the degree to which a product, device, service, or environment is accessible by as many people as possible.

2) Maintainability: In software engineering, maintainability is the ease with which a software product can be modified in order to:
- Correct defects
- Meet new requirements

New Services can be added in the project based on the user requirements just by adding the appropriate files to existing project using C# programming languages and etc. Since the programming is very simple, it is easier to find and make changes in defects in the code.

3) Scalability: System is capable of handling increased total throughput under an increased load when resources (typically hardware) are added. System can work normally under situations such as low bandwidth and large numbers of users.

4) Scalability: Portability is one of the key concepts of high-level programming. Portability is a feature to able to reuse the existing code instead of creating new code when moving software from an environment to another project can be executed under different operation conditions provided it meet its minimum configurations. Only system files and dependent assemblies would have to be configured in such a case.

C. Hardware and Software Requirements

1) Hardware Requirements:
- Processor: Processor above 2 GHz
- RAM: 8 GB
- Hard Disk: 128 GB
- GPU: Nvidia GTX 1050 Ti or above

2) Software Requirements:
- Operating system: Windows 10
- Front End: (Not decided)
- IDE: Visual Studio Code, Sublime Text 3, Jupyter Notebook
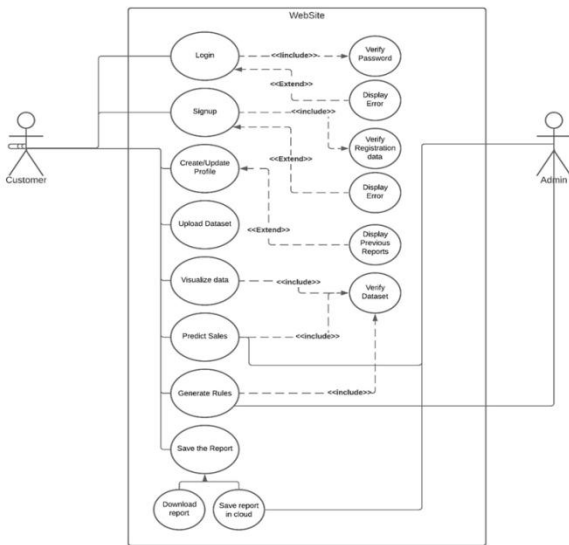- Server: Django, Dash

IV. DESIGN

A. Design diagrams

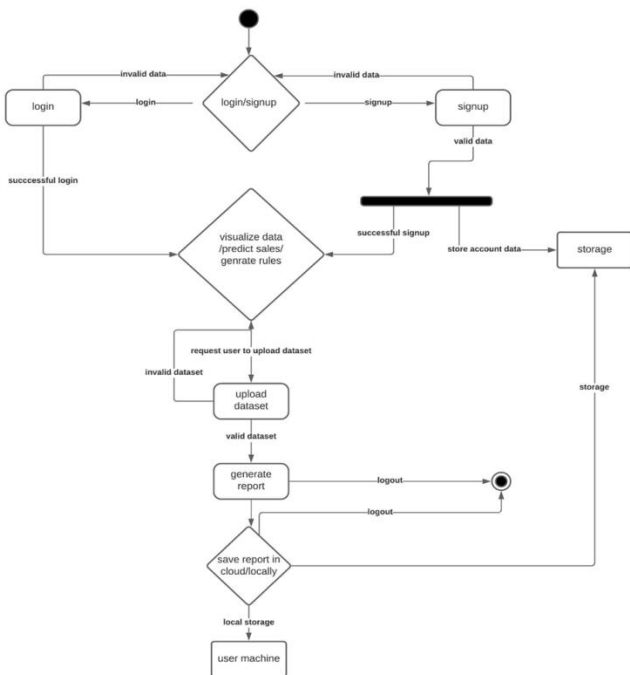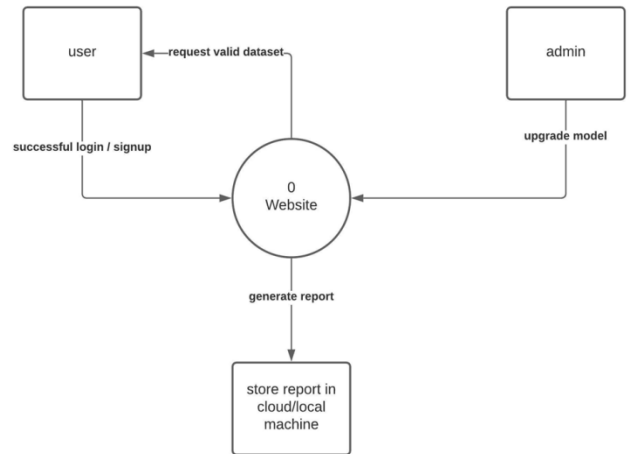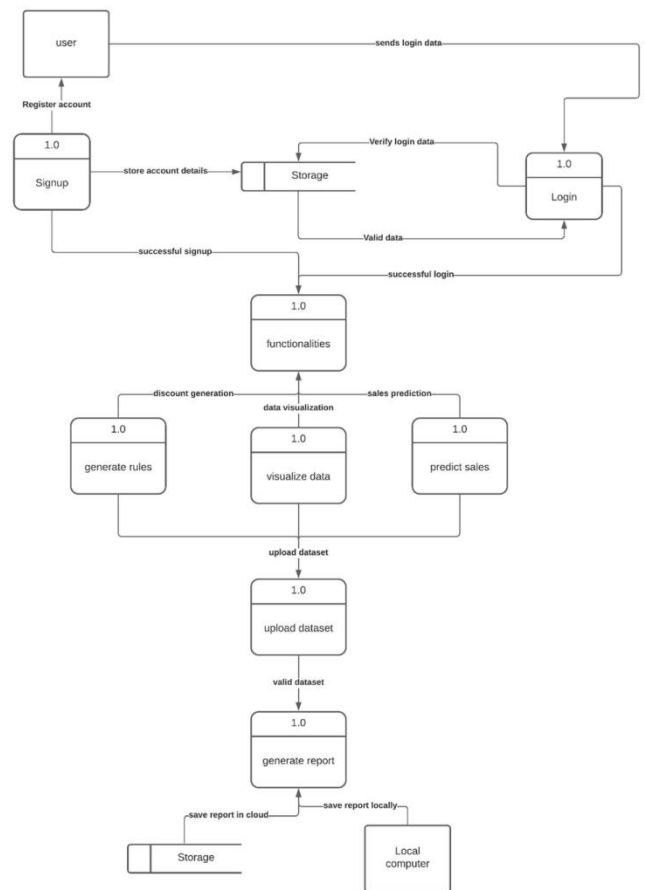Fig.6 Use Case Diagram

Fig.7 State Diagram

Fig.8 DFD Level 0

Fig.9 DFD Level 1

V. CONCLUSION

In this paper, various machine learning techniques have been utilized in this project: Facebook Prophet and fp-growth algorithms etc. on the sales dataset uploaded by the user. Both the techniques have

shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of recently Introduced machine learning techniques in the prediction of sales has yielded promising results without the effort of calculating the hyper parameters. It has led to the conclusion that it is possible to predict sales and market basket analysis, with more accuracy and efficiency using machine learning techniques. In the future, this project can be further improved by utilizing much bigger dataset than the one being utilized currently. This would help to increase the accuracy of our prediction models. Furthermore, other models of Machine learning could also be studied to check for the accuracy rate resulted by them.

VI. ACKNOWLEDGMENT

VII.   REFERENCES

[1]. Emir Žunić, Kemal Korjenić, Kerim Hodžić, and Dženana Đonko," APPLICATION OF FACEBOOK'S PROPHET ALGORITHM FOR SUCCESSFUL SALES FORECASTING BASED ON REAL-WORLD DATA" , Info Studio d.o.o. Sarajevo, Bosnia and Herzegovina Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina,Vol 12, No 2, April 2020

[2]. Aras, S., Deveci Kocakoç, İ. and Polat, C. (2017). Comparative study on retail sales forecasting between single and combination methods. J. Bus. Econ.   Manag. https://doi.org/10.3846/16111699.2017.1367324

[3]. Sean J. Taylor,Benjamin Lethamy,"Forecasting at Scale",Facebook, Menlo Park, California, United States,publ: 27 Sep 2017

[4]. Kiran Singh; Rakhi Wajgi,"Data analysis and visualization of sales data",Department of Computer Technology YCCE, Nagpur, India,06 October 2016

[5]. U. N. Umesh, Martin Kagan,"Data Visualization in Marketing",1 Professor of Marketing, Carson College of Marketing, Washington State University, Vancouver,10.15640/jmm.v3n2a4

[6]. Christian Borgelt," An Implementation of the FP-growth Algorithm"Paris Lodron University of Salzburg,January   2010, DOI:10.1145/1133905.1133907

[7]. Uma.N, Dr.Prashanth C S R," A Detailed Analysis of the various Frequent Itemset Mining Algorithms", Department of Computer Science and Engineering, New Horizon College of Engineering,   DOI: 10.5373/JARDCS/V12SP2/SP20201091

Classification of Tic-Tac-Toe Game Using Deep Neural Network

Sanjiv Sridhar[1], Shaumik Shetty[1]

[1]Department of Computer Science & Engineering, PES University, Bengaluru, Karnataka, India

ABSTRACT

Tic-Tac-Toe online game is a well-known two-player game that is played by using a 3x3 grid. The player who can put the own marker three-in-a-row will win this game. A machine learning classification model can predict the result of the tic-tac-toe game played between two players. This paper presents the use of a deep learning model to precisely classify the Tic-Tac-Toe game outcome. This study analyses several machine learning methods and deep neural network models for classifications of the tic-tac-toe game. The current study uses the online available UCI repository dataset for tic-tac-toe game classification. The study implements the predictive model and evaluates the performance using accuracy and loss as evaluation metrics. The experimental results prove that the deep neural network model can successfully predict the correct outcome of the game.

Keywords — Tic-tac-toe online game, Machine learning methods, Deep neural networks, Online board games, Predictive Modelling

I.   INTRODUCTION

The Tic-Tac-Toe board game is originally designed to play using paper and pencil and is also called a game of crosses and noughts. The two players need to select their markers, either Xs or Os, and put down their markers in a 3X3 grid. The winning rule for this game is that a player who has successfully put his/her marker horizontally in a row, vertically in a column, or diagonally will win the game. Fig. 1 depicts a game won by a player with marker X because of three Xs presents in the row sideways. As shown in Fig. 1, each square can be designated as marked with an X, an O, or left blank (b). However, there are 255,168 possible ways of playing tic-tac-toe board or paper game [1].

X	O	O

b	X	b
b	b	X

Fig. 1. Winning position of Tic-tac-toe game

The machine learning method has proved its capabilities to perform predicting modeling successfully. The paper examines whether the classification methods can successfully predict the Tic-Tac-Toe game outcomes. Researchers. However, the study focuses on finding a suitable predictive algorithm to classify the game dataset using several machine learning and deep neural network models. The supervised learning method is a classification technique used to build a predictive model from training data with the outcome variable. The data samples of the training dataset consist of several characteristics as input variables and an output label. The Tic-Tac-Toe board game is considered a binary classification problem. There are nine positional input

features and an output variable based on the marker's positions. Every feature holds an "X" an "O" or blank as b. The binary output label can be either "X" or "O." The rest of the content of this paper is presented as follows. Relevant existing work is analyzed in the second section. The third section explains the used methodology. Experimental findings are reported in the fourth section. The fifth section concludes and highlights future directions for this current work.

II. RELATED WORK

Several classification techniques are used to predict the outcome of the tic-tac-toe board game. Few studies are analysed here to identify the performance of the various predictive model using this game dataset.

The author evaluated seven machine learning models and three evaluation methods to classify the tic-tac-toe game dataset [2]. The experimental results showed that the 3-Nearest Neighbour classifier outperformed others model with an accuracy rate of 99%. The study investigated in [3] created artificial intelligence-based Tic-Tac-Toe 4x4 game using Minimax algorithm. This study adopted and modified the subset of rules using the best gameplay scenarios. The study predicted the precise moves and helped the user to win the game.

The authors compared and evaluated the predictive performance of several classification techniques using three different strategic board games datasets. WEKA tool was used to perform the experiments, and the Logistic Model Tree showed the better prediction results as 98.23% over other models [4]. This study proposed in [5] used both association rules and decision trees to identify the rules for Tic-Tac-Toe games. The experimental findings showed that the ID3 decision tree and association rule miner as Predictive Apriori algorithm identified 100% of game rules.

The study conducted in [6] classified the Tic-Tac-Toe game dataset using Decision Tree, Naïve Bayes, Random Forest, and Neural Network. Several data mining tools such as WEKA, Rapid Miner, and Orange were used to analyze the performance of these classifiers. The results demonstrated that the accuracy of neural networks in Rapid Miner was achieved as 98.20% than other data mining tools and other classifiers.

The author developed a model by training Artificial Neural Network to predict tic-tac-toe board games. The model analyzed the mathematical combination of the sequences using both the Gradient Descent Algorithm explicitly and the Elimination theory rules implicitly. The system produced solutions to solve each game stage to win the game or draw [7].

The above-analyzed studies showed that several machine learning algorithms were identified and used to classify the tic-tac-toe board game. However, the deep neural network can effectively extract the features of this dataset and predict the outcome precisely due to its training capacities [8,10]. Therefore, this paper focuses on evaluating deep neural networks along with machine learning-based classifiers.

III. METHODOLOGY

The dataset is retrieved from the UCI repository, a set of 958 possible sample moves of the tic-tac-toe board game. The dataset contains nine features representing the nine squares of a board for a tic-tac-toe game. The tenth attribute of the dataset represents the class label which denotes if the outcome of described board configuration is positive means wining or negative means losing the game for player X. However, there are 255,168 possible ways of playing tic-tac-toe board or paper game.

Each square can be designated as marked with an X, an O, or left blank (b) on the board as it is visible. The mapping of variables to physical squares is shown in Fig 1. Remember, the outcomes are positive or negative based on X winning. Fig. 2 portrays an example board endgame layout, followed by its dataset representation.

To use the Tic-Tac-Toe board game as input the deep neural network, the model translates these 3X3 boards into numerical vectors. The board of the Tic-Tac-Toe game, as illustrated in Fig.2, has nine cells as 1,2…,9 with either marked or empty with "Xs" or "Os."

1	2	3
4	5	6
7	8	9

Fig. 2. Positional attributes of the Tic-tac-toe game

The training dataset has a set of attributes f_i for this game as along with the output label as p_i as below in (1) and (2) :

$$f_i = \{f_1, f_2, …, f_9\} \qquad (1)$$

$$p_i = \{c_1, c_2\} \qquad (2)$$

The deep neural network finds the model from the data instances, where the Tic-tac-toe game with features f_i is denoted using nine positional attributes. The deep neural network classifier is trained to predict the correct outcome p_i as given in (3).

$$p_i = g(f_i) \qquad (3)$$

## A. Data Preparation of Training Deep neural Network Model

The proposed framework builds a **deep neural network** that learns the Tic-Tac-Toe board game. The model takes the board game state with nine features as provided with Table I as an input, and the predicted label is either positive or negative.

Sr. No	Features	Values
		b=Blank, O-Marker for player 1, X-Marker for player 2
1	top-left-square	{b, O, X}
2	top-middle-square	{b, O, X}
3	top-right-square	{b, O, X}
4	middle-left-square	{b, O, X}
5	middle-middle-square	{b, O, X}
6	middle-right-square	{b, O, X}
7	bottom-left-square	{b, O, X}
8	bottom-middle-square:	{b, O, X}
9	bottom-right-square:	{b, O, X}
10	Class	{positive, negative}

Data pre-processing and transformations are performed on the dataset before feeding the data to the deep neural network [9]. Fig 3. presents the heatmap of correlation among the Tic-tac-toe board game. The model encodes the categorical features into numeric values using Label Encoder class from Scikit-learn pre-processing module. It converts from {b, O, X} to {0, 1, 2} for every variable. Similarly, the categorical class variable is converted {negative, positive} to {0, 1}.

TABLE I

The Description of The Tic-Tac-Toe Game Dataset
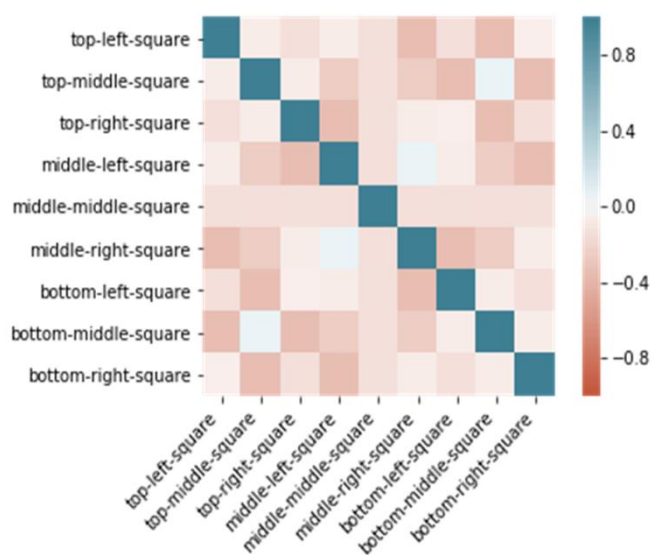
**Fig. 3.** Correlational matrix for attributes of Tic-tac-toe game

In the next step, the model uses Scikit-learn's train/test split function to divide the dataset into training and testing sets. Table II provides the parameter details to construct a deep neural network for the Tic-tac-toe board game. The model is built using TensorFlow 2.0 and Keras to perform the tic-tac-toe board game binary classification.

TABLE II

The Description of Parameters Used To Build The Model

| Sr.No	Parameters/Hyperparameters		Value
1	Input neurons		9
2	Hidden Layers		5
3	Hidden Layers Neurons		10
4	Activation Function	Hidden Layers	Relu
		Output layer	Sigmoid
5	Optimizer		Adam
6	Loss Function		binary-cross-entropy
7	Weight initialization		Random
8	Learning rate		0.001
9	epoch		100

IV. RESULTS AND DISCUSSIONS

The Tic-tac-toe dataset is trained and evaluated using a Decision tree, K-Nearest Neighbours with two neighbors (KNN), Support Vector Machines (SVM), Artificial Neural network (ANN) with one hidden layer, and Deep Neural Network (DNN). The experimental findings, as demonstrated in Fig. 4, show that the maximum accuracy for the trained deep neural network with five hidden layers with nine neurons in each hidden layer is reached 98.90%. The results prove that the DNN model predicts the game outcomes better than other base machine learning classifiers.
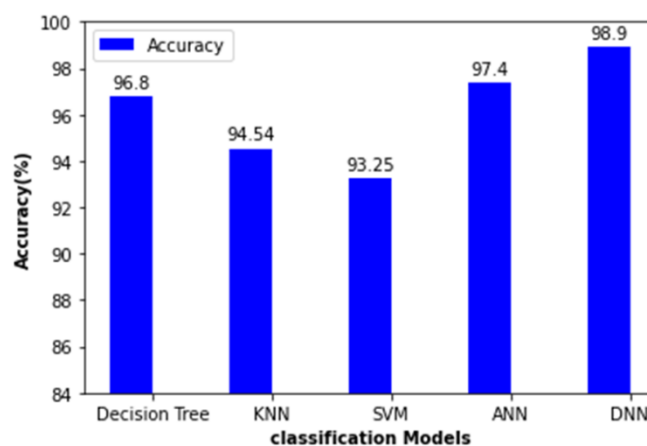
Fig. 4 Comparative analysis of different predictive models

The model attempts the different optimizers instead of Adam optimizer, different neuron sizes in each hidden layer, different number of layers for the hidden layer, and several learning rates. However, the resulting trained DNN model did not result in better accuracy, loss, or correctly predicted test data instances beyond what we have reported above. Fig. 4 and Fig. 5 have shown the plots for accuracy and loss achieved by the Deep Neural Network over the 100 epochs.
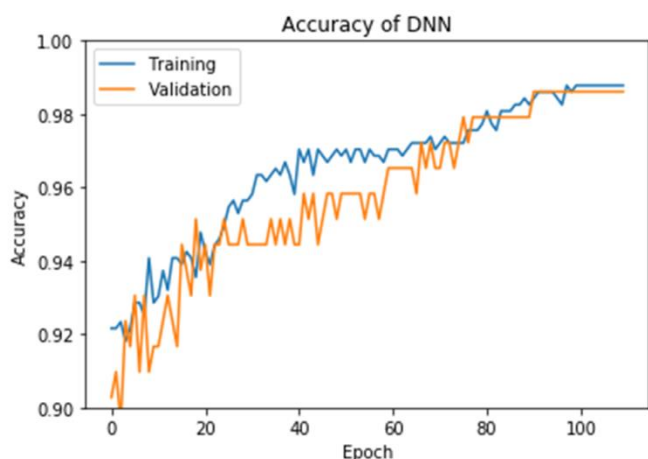
Fig. 4 Accuracy plot for DNN

From the accuracy plot, as given in Fig. 4, we can infer that the DNN model has not yet over-fitted the training dataset and shows comparable better results for the Tic-tac-toe dataset.
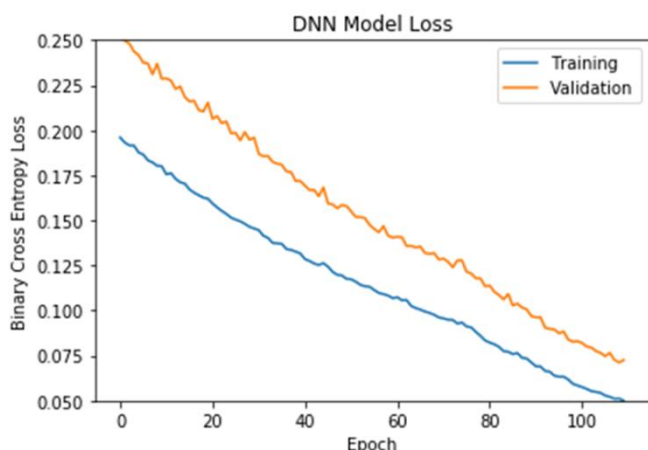
Fig. 5 Model Loss plot for DNN

From the loss plot, as given in Fig. 5, we can infer that the DNN model has comparatively better performance with both training and validation datasets.

V. CONCLUSIONS

The paper presents the Deep Neural Network (DNN) to predict the Tic-tac-toe board game by training the model with precise parameters and hyperparameters. The study has contributed towards deep learning research by comparing and analyzing the

effectiveness of base machine learning classifiers and Deep Neural Networks. This study identifies the applicability of these different classifiers with their flaws and benefits in implementing a system to predict the Tic-tac-toe board game. In the future, we can apply the various rule-based algorithms or reinforcement machine learning methods to improve the winning possibilities of the game.

VI. REFERENCES

[1]. S. Schaeffer. (2002). Tic-Tac-Toe (Naughts and Crosses, Cheese and Crackers, etc.). Mathematical Recreations. [Online]. Available: http://www.mathrec.org/old/2002jan/solutions.html

[2]. C-H. Chou, Using Tic-Tac-Toe for Learning Data Mining Classifications and Evaluations. International Journal of Information and Education Technology. 437-441. 10.7763/IJIET.2013.V3.314, 2013

[3]. R., A. Makame, H. Nadhira & R. Abdul-Rahimsidiq. Tic-Tac-Toe Learning Using Artificial Neural Networks. International Journal of Engineering and Information Systems, Vol. 3. : 1-6, 2020.

[4]. M. Gulzar, A. Ali and B. Naqvi, Performance Evaluation and Comparison of Classification Techniques for Outcome Estimation in Strategic Board Games, International Journal of Computer Science and Network Security, Vol.18 No.7, 2018

[5]. C-H. Chou, Using Data Mining Techniques for Game Rule Extraction, Issues in Information Systems Volume 20, Issue 3, pp. 117-127, https://doi.org/10.48009/3_iis_2019_117-127, 2019

[6]. J.S. Ravi, G. Mohan Seshu, and M. S. K. Vamsi Varma, Relative Exploration of Classification Techniques of Machine Learning on Various Data Mining Tools, Journal of Xi'an University of Architecture & Technology, Vol. XII, Issue IV, pp-5150-5156, 2020.

[7]. M. A.Dalffa, ANN for Tic-Tac-Toe Learning, International Journal of Engineering and Information Systems (ISSN: 2000-000X Vol. 3 Issue 2, pp- 9-17, 2019.

[8]. L. Munkhdalai, T. Munkhdalai, K-H. Ryu, GEV-NN: A deep neural network architecture for class imbalance problem in binary classification, Knowledge-Based Systems, Vol 194, 105534, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2020.105534, 2020

[9]. Z. Chen, M Pang, Z. Zhao, S. Li, R. Miao, Y. Zhang, X. Feng, X. Feng, Y. Zhang, M. Duan, L. Huang, and F. Zhou, Feature selection may improve deep neural networks for the bioinformatics problems, Bioinformatics, Vol 36, Issue 5, pp. 1542–1552, https://doi.org/10.1093/bioinformatics/btz763, March 2020.

[10]. B. Smrithi, P. Anil1, P. K. Devi, R. M. Midhuna, S. Kumar Nanthana, and S. Sreejith, Classification Framework Using a Customized Feed Forward Neural Network, Journal of Innovation in Information Technology, Vol. 4(1), pp-28-33, 2020

A Survey on Detection of Phishing Websites using Machine Learning

Revati Pote[1], Anjali Potdar[1], Shubhangi Sapkale[1], Manorama Jadhav[1], Deepali Ujalambkar[1]

[1]Department of Computer Engineering, Sri Savitribai Phule Pune University, Ganeshkhind, All India Shri

Shivaji Memorial Society's College of Engineering, Pune Kennedy Road, near RTO Office, Sangamwadi,

Shivajinagar, Pune, Maharashtra, India

ABSTRACT

Phishing internet sites contents and internet-predicated consummately data includes varied hints. The victim's personal and sensitive record is obtained by phishing sites which lead them to surf a phishing internet site that resembles a valid internet site, that's one of the illegal assaults triumphing with inside the cyber world. The proposed a brilliant version for detecting phishing internet pages primarily predicated on Extreme Learning Machine. Types of internet pages are one of a kind in phrases in their features. Hence, we require to utilize a web page feature set to preserve any phishing assault. A Machine Learning approach is implemented to resist these attacks.

The projected technique for importing phishing dataset, legitimate URLs from the database, and also data that is obtained are pre-processed. Phishing website detection is performed on four classes of URL features: domain, address, abnormal based, HTML, JavaScript features. With the aid of processed data URL features are extracted also values for URL attribute are generated. URL analysis is performed by ML techniques that calculates the threshold value as well as range value for URL attributes. The objective of this project is to implement an ELM classification for several features and some phishing sites within the database.

Keywords— Browser extensions, Extreme Learning Machine (ELM), (SVM) Support Vector Machine ,URL Phishing Websites.

I. INTRODUCTION

Phishing is the sham plan to acquire confidential facts which include username, password and credit card info, customarily for malignant functioning by dissimulating itself as a sincere entity in an electronic interaction [1]. Phishing has become a worrisome concept for safety researchers nowadays as it's not arduous to engender a faux internet site that looks similar to a legitimate website. It is easy for specialists to recognize a faux web sites however it is

tough for all users to distinguish between them and such users emerge as sufferers of phishing attacks. The assailant's main motive is to thieve financial institution account credentials. US companies lose $ 2 billion annually as their customers end up being sufferers of phishing. [3]

The 3rd Microsoft Computing Safer Index Study published in February 2014 reported that the annual worldwide impact of phishing could be near to or more than $5 billion. One of reasons why these attacks are prospering is due to lack of a person's

apprehension. Since the phishing attack takes undue advantage of the vulnerable data of the users, it's far very hard to mitigate them, however it is very vital to amend the phishing detection strategies. [3]

In this assailment, Phisher makes a faux internet web page by replicating contents of the valid web page, in order that a person can't differentiate among phishing and legal sites. Social engineering schemes prey on unwary sufferers by bamboozling them into believing they're managing a trusted, valid party, while using misleading e-mail addresses and e-mail messages. [1]

The overall approach to ascertain phishing websites via updating blacklisted URLs, IP to the antivirus database which is likewise known as 'blacklist' method. To stay away from blacklists, assailers use ingenious strategies to illude customers into editing the URL to seem legitimate thru obfuscation and various simple techniques inclusive of: fast-flux, proxies are generated automatically to host a web page, etc. [3]

It is viable to utilize ML to get acquainted with and develop brilliant data outputs. The system objects to discover this concept by exhibiting a use-case of detecting phishing websites. [13]

ML methods can moreover be implemented in information safety, particularly for its application development. Prediction, optimization, decision-making, classification and large benefits may be given to the one in charge for information security (3)

The common phishing attacks may be executed through e-mail phishing scams and spear phishing thus client must be privy to the results and have to know no longer offer their one hundred percent trust on any unauthorized security application. The disadvantage of existing approach can be overcome using ML. [3]

This is an area of artificial intelligence that has the potential to learn without explicit programming. Various machine learning techniques, unsupervised learning, and reinforcement learning are supervised. The types of machine learning techniques are:

A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning

A. Machine Learning Algorithm

(ELM) Extreme learning machine: the extreme learning machine (ELM) is an Artificial Neural network (ANN) model with one hidden layer. so as for ANN to ensure advanced learning, parameters equal to threshold value, weight, and activation perform have to contain values suitable to the data system to be modelled. In gradient-based learning approaches, all of those parameters are modified again and again to relevant values.

B. Random Forest Algorithm

Random forest algorithm: Random forest (RF) is a set learning regression and category technique suitable for handling problems related to the grouping of data into classes. In RF, prediction is achieved the usage of decision trees. during the training phase, a few decision trees are built (defined via the programmer) which are then used for class prediction; this is performed via considering the graded classes of all individual trees and the class with the highest grade is considered the output. [10]

C. Support Vector Machine

This technique is utilized in medical for diagnosis of diseases, textual content recognition, for class of image and within the other fields. This will partition the data into classes the use of fixed rule, quadratic equation and statistic. Separating hyper plane is used for the binary classification of the data and minimizes the space of the margin on the basis of kernel characteristic. This technique is used to find the best solution of the problem. This technique is fails in analysing the huge data. [2]

II. LITERATURE REVIEW

The device relies on a machine learning method, specifically supervised leaning. Here the Random forest method is selected for its smart ranking overall performance. The aim is on tracking the best executing classifier with the aid of examining the features of phishing website and determining the superior combination to train the classifier. Thus, the paper has accuracy 98.8% and the no of various features used are 26. [9]

This analysis proposes a framework that makes use machine learning systems to overcome the problem of spam. The framework has been modeled at the Azure level and moreover department of the e-mail servers has been examined. Developing a phishing detection model using varied data mining techniques to improve the accuracy of phishing detection and a feature selection methodology is additionally accustomed to surge the precision of the classification model by culling the most efficacious feature and finding the first-rate result.

Feature hashing utilizes Vowpal Wabbit which is a fast ML framework, with the aid of hash functions hashes feature words in n memory indexes. The paper presents 2-class logistic regression, , neural network ,boosted decision tree and SVM to distinguish any unsolicited approach. [2]

A real-time technical approach is projected with a purpose to effectively protect a consumer from client-side phishing attacks. Just one effectual feature of 'hyperlinks present in webpage' is used for detecting the attacks. Google public DNS is compared with IP address of the dubious sites to determine a DNS intrusion of devices. [4]

This paper presents ways for sleuthing phishing internet sites via analyzing numerous options of benign and phishing URLs by using machine learning strategies. Different methods are applied for detection of phished net-sites which supports lexical features, host and page consequentiality properties.

Examination of sundry data processing algorithms for analysis of the features is carried out so as to urge a higher understanding of the structure of URLs that leads to attack. The first-class-tuned parameters are helpful in deciding on the perfect ML algorithmic rule for separating the illegitimate sites from benign websites. [5]

The paper proposes Agile Unified Process (AUP) lifecycle to diminish the development stage. Admin has the authority to distinguish between blacklisted and whitelisted URLs, once these sites are inserted, he may edit, modify and delete it. For the users' convenience different color backgrounds are used to categorize phished or blacklisted URLs. The non-blacklisted URL will be opened when clicked on the link. The proposed machine identifies and selects to cope with the complexity of monitoring requisites for any contemporary scenario. This software focuses on the widespread level of its capability, features that exhibits inside the monitoring phase. [6]

The methodology implements an agent-predicted design and ML classifier for dealing with various forms of phishing attacks. Distributed internet requires the utilization of multi-agents which transmits via peer-to-peer method. The paper confers a layered multiple-agent system for distinguishing and resolving net based phishing attacks. The role of multi-agents is to extract URL, detect script and phished URL, block. The approach is successful to apprehend and study in step with the environmental adjustments. [8]

The paper projects a deep learning model primarily drew on 1D CNN for phishing detection. The system analyzes a standard dataset which consists of 4,898 instances for phishing sites and 6,157 instances for legitimate sites. This model extraordinarily surpasses other favored ML classifiers who have been evaluated on the similar dataset. The final results stipulate that compared to various model the CNN based approach gives the most accurate outcome and detects new phished websites too. [7]

III. ANALYSIS AND RELATED WORK

There are numerous processes to securing URL phishing assaults. These processes could be labeled supported the real mechanism used. We will be inclined to analyze numerous ways of phishing detection. For the duration of this paragraph, we stated the techniques that we examined. [2] This paper presents phishing detection model by utilizing victimization numerous facts processing techniques & characteristic desire method (VowPal Wabbit) are wont to magnification the precision of relegation model by denotes of culling excellent characteristic & result. [4] The method to bulwark in opposition to phishing assault the utilization of white-list of valid accessed by character utilizer, checking legitimacy the utilization of link functions, descries phishing attacks for DNS poising, embedded objects, 0-hour attack. [7] This paper proposes a deep learning model supported 1D CNN for the detection of phishing websites. The outcomes designate that projected CNN predicated model will be wont to discover incipient, antecedently unseen phishing web sites as it should be. [8] They have enforced a multi- agent-predicated layout and MLl classifier for detective work and rectifying net phishing attacks. [9] This system, gives an intelligent system supported a ML approach for detecting phishing web websites, a similarly practicality is there may be an extension to a cyberspace browser that notifies person as soon as phishing internet site is detected. [6] The system fosters numerous options cherish capturing blacklisted URLs from the browser directly to verify the validity of the cyber world site, notifying utilizer on blacklisted websites while they're endeavoring to access through pop-up, and moreover notifying via e mail.

At some point of this segment, the perspicacious model is predicated on machine learning strategies to ascertain phishing internet pages. In proposed contrivance, imports a dataset of phishing and legitimate information from the database. Then the imported dataset is preprocessed. The detection of phishing web sites is accomplished by four edifications of deal with features: domain based, address based, abnormal based and HTML, JavaScript functions. URL is the primary detail to prognosticate a website to determine whether or not it is phishing or not. a few capabilities are concerned at the same time as URL is processed like digit matter within the URL, overall period of URL, checking whether the URL is hijacked or no longer, checking whether it includes a legitimate emblem call or not, quantity of sub domains in URL. The reason of phishing domain detection is detecting phishing domain denominations. A few subsidiary domain-primarily predicated capabilities like its domain name or its IP address in blacklists of apperception offerings? How many days exceeded because the domain turned into registered? Is the registrant's name hidden?

IV. THE ARCHITECTURE

At present, the general public of the populace has been illuded into giving their non-public statistics to a hacker or a phisher without even descrying it. with a purpose to expand this application i.e., phishing detection, a technique ought to be described and defined, the proposed approach that imports a phishing information set and legitimate URLs constitutes the dataset whilst the imported records are pre-processed. This mission may be carried out utilizing machine learning. The development duration and the flexible method are betokened within the figure 1.
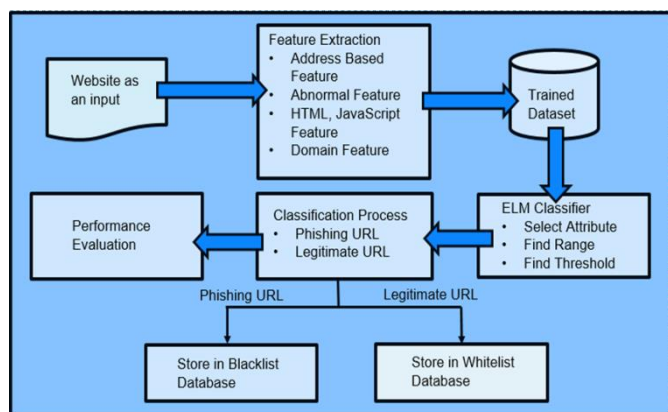
Fig.1. System Architecture

In the figure the architecture of Detection of Phishing will be explained. The first task of the user will be to add the extension on chrome window which will be followed by entering an URL or website. For detection the first task is Feature Extraction, following are its types:

- Address based feature extraction
- Abnormal feature extraction
- HTML/JavaScript feature extraction
- Domain feature extraction

Attribute values are calculated utilising feature Extraction. The phishing attribute function is extracted for every URL to ascertain if a URL is phishing or valid. The URL_of_Anchor tag attribute is culled to find the overlap values: it is the sum of the value of the culled attribute this is combined with any other attribute.

For instance, if '@' symbol is present in the given URL then it takes '1' as input otherwise it takes '0'. Similarly, if another parameter is URL length and its length is 51 then input is considered to be '0' while if URL length is between 51-75 then it is considered as '1', finally if it is above 75 then it takes as '-1'.

Feature Extraction is performed using Trained Dataset. The proposed system uses dataset from the Kaggle.com. In Trained Dataset 28 patterns/parameters are used which are '@', 'URL length', 'dash in line (-)', 'dot in line(.)', etc. 28 parameters are used because of Time Complexity. The

parameters and time complexity are inversely proportional. As the increase in number of parameters causes a rise in the in the time complexity. The analysis of the URL is accomplished with the aid of machine learning which calculates the range value and threshold value for URL attributes. Out of 4 algorithm, 1 will be chosen on the basis of best accuracy and those algorithms are ELM, SVM, RF and LR. These comparisons are then followed by the classification process.

The aspect value for each URL is calculated making use of the corresponding set of aspect values $\{-1, 0, 1\}$. aspect X that URL_of_Anchor tag cost and aspect Y that is Prefix_Suffix price. both the URL_of_Anchor tag aspect and the prefix suffix withal have an interrelated fee and that calls for to be calculated to find the variety threshold price. For example, if URL output turns out to be '0102' then it can be justified to be a legitimate URL while if a URL output is '00010' then it can be classified as suspicious URL finally if URL output turns out as '11100' then it can be said that it is a Phished URL. If the URL is a phished one then it is stored into the Blacklist database and if an URL is legitimate then it is stored into the Whitelist database. Once classification is completed, the result is displayed to user which states whether the entered URL is phished or legitimate. If URL is phished URL, then a pop-up window alerts the user by display the message, 'URL is phished don't go ahead...!'

V. ALGORITHM AND SEQUENCE FLOW

The SVM classifier is that the maximum generally used machine learning classifier to set off the best line between two lessons.

Logistic regression is used for the classification problems and it is a predictive analysis algorithm.

Random forest only searches within randomly selected predictors for the best possible split; has good performance in classification.

Step 1 : Start by using deciding on random samples from a given dataset.

Step 2 : Then

this formulation will build a decision tree for each pattern. Then it's going to get the outcomes the outcomes of the prediction of every decision tree.

Step 3 : At some stage in this step, the vote are carried out for each predicted result.

Step 4 : Subsequently, pick the

very exceptional rated forecast result due to the fact the final forecast end result.

In ELM studying approaches, unlike ANN that renews its parameters as supported gradients, the input weights are every which manner hand-picked whereas the o / p weights are analytically calculated.

Step 1 : Visit internet site or an internet web page

Step 2 : Check the thirty enter attributes supported characteristics and their policies

Step 3 : Grouping samples to the dataset.

Step 4 : Indiscriminately chosen 90% training sample s and 10% trying out samples of the dataset.

Step 5 : Classification by means of the use of ELM.

5.1: At random generate hidden nodes parameters and assign hidden nodes randomly.

5.2: Calculate the output matrix of the hidden layer.

5.3: Calculate the output weight matrix.

Step 6 : Prediction for phishing or legitimate.

VI. PROPOSED METHODOLOGY

The planned method for importing phishing information sets and valid computer address from the records and therefore the imported facts is preprocessed. Phishing internet web page detection is done supported 4 training of URL functionality: domain-based, abnormal based, address-based, and HTML, JavaScript features. These URL traits are extracted with processed data and values are generated for every URL characteristic. The analysis

of the URL is finished using a machine learning method that calculates the range cost and therefore the edge value for the attributes of the URL. It's then categorized into phishing and official URL. The characteristic values are calculated by extracting traits from phishing websites and are wont to decide the variety value and the edge cost.

VII.   CONCLUSIONS

Systems ranging from records entry to scientific applications is created via web sites. The input statistics may be processed; processed information may be acquired as output. In recent times, web sites are applied in numerous fields like medical, technical, business, training, economics, and so forth. Because of this extensive use, it could also be used as a device by using hackers for malicious purposes. A malicious item seems to be a phishing assault.

Many analysis contributions display totally unique te chniques, procedures to locate phishing URLs and those methodologies have additionally been applied. The aim of the equipment is to shape a category for the determination of 1 of the styles of attacks that cyber threats decision phishing. The system informs the consumer of phishing URLs by means of suggesting benign URLs even before it's found out on such web sites that finally ends up averting a phishing assault. For this reason, the intense learning gadget are going to be used. In the course of this look at, we're going to use a dataset from the UCI internet site.

VIII. REFERENCES

[1]. OzaPranali P, Deepak Upadhyay, Review on Phishing Sites Detection Techniques, IJERT, ISSN: 2278-0181, 04, April-2020

[2]. Meenu, Sunilagodara, Phishing Detection using Machine Learning Techniques, IJEAT, ISSN: 2249 – 8958, 2, December, 2019

[3]. Sandeep Kumar Satapathy, Shruti Mishra, Pradeep Kumar Mallick, LavanyaBadiginchala, Ravali Reddy Gudur, Siri ChandanaGuttha, IJITEE, ISSN: 2278-3075, June 2019

[4]. Ankit Kumar Jain and B.B.Gupta EURASIP Journal on Information Security (2016) 2016:9

[5]. Joby James, Sandhya L., Ciza Thomas, Detection of Phishing URLs Using Machine Learning Techniques, 2013 International Conference on Control Communication and Computing (ICCC), December 2013

[6]. Mohammed HazimAlkawaz, Stephanie Joanne Steven, Asif Iqbal Hajamydeen, Detecting Phishing Website Using Machine Learning, 2020 16th IEEE International Colloquium on Signal Processing & its Applications, 28-29 Feb. 2020

[7]. Suleiman Y. Yerima, Mohammed K. Alzaylaee, High Accuracy Phishing Detection Based on Convolutional Neural Networks, IEEE Xplore

[8]. Megha N, K.R.RemeshBabu, Elizabeth Sherly, An Intelligent System for Phishing Attack Detection and Prevention, IEEE Xplore ISBN: 978-1-7281-1261-9, 2019 IEEE

[9]. Amani Alswailem, BashayrAlabdullah, Norah Alrumayh, Dr.AramAlsedrani, Detecting Phishing Websites UsingMachine Learning 978-1-7281-0108-8/19/ 2019 IEEE

[10]. https://www.hindawi.com/journals/jam/2014/425731/ (random forest)

[11]. https://pdfs.semanticscholar.org/41ca/257920b5b5e6c1cf4f4417bb85ac5a875935.pdf

[12]. https://archive.ics.uci.edu/ml/index.php

[13]. https://www.google.com/

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com
Email: info@technoscienceacademy.com

Third National Conference on Advancements in Computer Science and Engineering

Organised by
Department of Computer Science and Engineering,
New Horizon College of Engineering
In Association with
VTU, Belagavi & Computer Society of India