INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Future Prognostication of Diabetes using Exploratory Data Analysis

Miss Apurva Makarand Somvanshi[1], Master Abhijeet Aher[2], Master Jaspreet Renu[3], Master Bhavesh Baliram Lande[4], Dr. Amit Gadekar[5]

[1, 2, 3, 4]*Student,* [5]*Professor, Computer Engineering, Sandip Institute of Technology and Research Centre,* Nasik, Maharashtra

Abstract: Diabetes is one of the leading causes of death in the world. It is also the originator of a variety of enemy problems: heart failure, blindness, urinary tract infections etc. In that case the patient needs to visit a diagnostic centre, to get his or her reports after consultation. For each season they have to invest their time and money. But with the advent of Machine Learning Methods, we have found the flexibility to seek a solution to the current problem, we now have a detailed analysis of advanced system abuses that have the ability to predict whether a patient has polygenic disease or not. In addition, predicting illness early on ends up giving patients before it starts is important. Withdrawal of information is free to delete invisible data from a wealth of information associated with diabetes. The purpose of this analysis is to develop a system that can predict a patient's diabetes level with better accuracy. Model development is based on classification methods such as Decision Tree, ANN, Naive Bayes and SVM algorithms. At Decision Tree, the models offer 85% specification, to Naive Bayes 77% and 77.3% of Support Vector Machine. The results show significant accuracy of methods.
Keywords: Machine Learning, Support vector machine, Artificial Neural Network, Decision Tree, Naive Bayes, Data Mining.

I. INTRODUCTION

Diabetes has already become a major threat to human health. Many fall victim to it and are unable to get out of it despite the fact that they are working to prevent it from growing. Cloud Computing and Internet of Things (IoT) are two tools that play a very important role in modern life in terms of many aspects and objectives including monitoring the health care of patients and the elderly community. Diabetes Health Care Monitoring Services are very important these days because even remote health care monitoring because going to hospitals and standing in line is an ineffective concept of monitoring patients. If a patient has diabetes that is very incurable and spends his or her time standing in line anything dangerous can happen to him or her at any time. Therefore, this paper came up with clever sensors and different machine learning techniques like xgboost algorithm, a random forest. Diabetes can act as a cure for other ailments such as heart disease, kidney damage and loss of vision. It can use a variety of machine learning techniques such as vector support, line layout, tree deck, xgboost and random forest with its help that can easily find the full functionality and accuracy of predicting whether a person will suffer from diabetes or not. There are many different traditional methods that are completely different from the software methods that can detect diabetes and predict pre-diabetes conditions. Diabetes is a result of high blood glucose levels. There is a utility system available using the train test split and K fold cross verification using the Sklit learning method.

There has been a dramatic increase in the number of people with diabetes over the past decade. The current lifestyle is the main reason for the development of diabetes. With the current medical diagnostic method, there can be three types of errors-

A. A false type in which a patient is actually already a diabetic but the test results tell that person does not have diabetes.
B. False type. In this case, the patient is not really a diabetic but test reports say that a diabetic.
C. The third type is the default type where the system will not be able to diagnose a given case. This is due to the removal of insufficient information from previous data, the given patient can be predicted in the uninstalled type.

However, in reality, the patient should predict whether they are in the diabetes or non-diabetes category. Such diagnoses can lead to unnecessary or non-existent treatment. To avoid or reduce the severity of the impact, there is a need to develop a system that uses machine learning techniques and data mining methods that will provide accurate results and reduce human effort.

II. LITERATURE SURVEY

Veena Vijayan V. And Anjali C has discussed, the diabetes disease produced by rise of sugar level in the plasma. Various computerized information systems were outlined utilizing classifiers for anticipating and diagnosing diabetes using decision tree, SVM, Naive Bayes and ANN algorithms [1]. P. Suresh Kumar and V. Umatejaswi has presented the algorithms like Decision Tree, SVM, Naive Bayes for identifying diabetes using data mining techniques [2]. Ridam Pal, Dr.Jayanta Poray and Mainak Sen has presented the Diabatic Retinopathy (DR) which is one of the leading cause of sight inefficiency for diabetic patients. In which they reviewed the performance of a set of machine learning algorithms and verify their performance for a particular data set [3]. Dr. M. Renuka Devi and J. Maria Shyla has discussed about the analysis of various skills of mining to guess diabetes using Naive Bayes, Random forest, Decision Tree and J48 algorithms [5]. Rahul Joshi and Minyechil Alehegn has discussed the ML techniques which are used to guess the datasets at an initial phase to save the life. Using KNN and Naive Bayes algorithm [6]. Zhilbert Tafa and Nerxhivane Pervetica has discussed the result of algorithms that are implemented in order to progress the diagnosis reliability [7]. Prof. Dhomse Kanchan B. and Mr. Mahale Kishor M. has discussed the study of Machine Learning Algorithms such as Support Vector Machine, Naïve Bayes, Decision Tree, and PCA for Special Disease Prediction using Principal of Component Analysis [10]. Beer Lambert Law is very much useful to find and calculate the oxylemetric pulse rate of the human body and to determine the behavior during chronic diabetes situation. The model system was built to calculate the oxymetric pulse rate of the human body successfully using a particular wireless sensor named ZigBee. But it has the biggest disadvantage regarding cost factors as it is highly costly and expensive and the pulse oxymetric analysis is not allowed on a daily basis activity and not widely used. So, such kinds of problem are arising and proper actions were also taken to handle it remotely. Now many applications are looking for corresponding wearable pulse oxymetric devices to enhance the overall throughput and system efficiency. All the wireless devices deal with the learning techniques such as random forest, decision trees, xgboosting gradient and smart cities have a point to point and another peer-to-peer interconnection for client services and maintenance purposes. This paper has to focus on the development of a system that is both output resilient and ecofriendly because in no way we can harm the nature in the name of healthcare monitoring of the patients and other clients to determine the process of diabetes prediction. This research paper will be collecting various data as per the latest records and gain experience from them to deal in the upcoming future prospects widely. If we try to focus on the data mining field then there, so our paper needs to deal with the extraction of hidden data from the huge collection of the Pima Indian datasets of diabetes.

The involvement of machine technical learning algorithms and smart medical sensors in the system produces a huge impact over the world. Another important motivation regarding it is to motivate people to take their daily life routine through physical activities such as calisthenics, body exercises and other schemes such as workouts etc. Comparatively an ambiguous health related solution is very much easier to obtain on an android based sensor network. The nature of the wireless sensors is totally noninvasive to be successfully getting interconnected with the mobile devices and computer monitoring system standards database widely because it is totally dependent on the learning purpose techniques in the side of data mining. The widely used health gear box for assigning the physiological parameters to deal with the signal standards purposefully. The set up totally is set like the above figure. Our paper will be sidely using a display to show up the contents and health status parameters in a digital manner. Then we may collect all the related holdings and sample details of the clients successfully. The indirect connection exhibits between the microcontroller and the ADC network. The all-related parameters regarding the patients are structurally stored inside the memory storage. In another way we can use an amplifier to amplify and further strengthen the parameters such as patient heartbeat, pulse rate, temperature etc. so that a proper strong digital signal is obtained for further hardware and software coordination. The use of calibration set up is mainly due to the fact of setting the surrounding pressure to an appropriate threshold value for a proper calculation scheme. The main focus is the increase in accuracy of the overall system and eliminating the chance of an error that is the main exact role of the calibrated system.

There is a particular reason why we are using a network, actually ADC is an analog to digital signal converter , since computers take input only in the form of digital signals that's why the paper needs to convert each and every analog signal input into corresponding digital signal to make it easy for the computer to analyses it for processing. Afterwards this review need to safeguard the information and data collected and prevent it from getting into wrong hands that is we must ensure that the parameters are only accessible to the authorized users and not to any outsiders and corrupted unauthorized. System users due to which this paper need holding sample collections that facilitates this task effectively with resilience.

This system widely increases the formal standard of the remote monitoring of server hosts. There is an additional alarm circuit whose purpose is to raise an alarm to the ambulance in case any abnormality is found in the health parameters in the cloud infrastructural database.

In this way it makes the caretakers alert timely without causing much interference while they are doing their work. One can imagine how much waste of time can occur if we keep on physically visiting the hospitals every day for the one same purpose all along the days. The end step is the most vital among all others that is the gain control which deals with the overall controlled mechanism preventing it from getting deviated to other path and mainly maintaining the system to remain in the main frame procedure and helps in keeping it specialized stable for future prospects.

Not only in the current and ongoing generations require remote healthcare but the oncoming and almost all of the future generations are totally going to be dependent on the digital smart technology to take necessary steps regarding health. One of the major key factors is balanced diet that is mainly lacking from the humans now a days. The system will totally manipulate over the fact that will make people understand what balanced diet is and how it can severely increase their level of health to a vast extent. People will seriously take conclusions and motivation from this research paper and try to take their time regarding the study of many research paper regarding the same task of healthcare related schemes.

Let us talk about the Wi-Fi module that exists in this system. The Wi-Fi module used in ESPR8266 that is so far the most effective version and cost-effective WIFI module used so far. The entire module structured with the ESPR8266 Wi-Fi rater is connected through the power supply for the origination of electrical supply as input power supply module. Both the network node and the radio links are completely synchronized with each other.

There exists a particular cloud structure known as the THINGSPEAK where we get the pie charts, tabular structures and the corresponding graphs of various sensor module outputs. On the other hand, it is also possible to connect multiple versions of MDP protocols to the vast techniques to enhance the input mechanism.

III. PROPOSED METHODOLOGY

As we saw that diabetes is non-curable disease but can be brought under control only if the patient get prior intimation . In this proposed system we are going to collect the data of patients with various attributes like: blood group, blood contents, eating habits and many more.

The data will be then trained using different tools. Using various algorithm like SVM, naïve bayes, decision tree, ANN we will try to extract features and get optimal analysis.

As more and more data sets will undergo the module. It will be trained to give best and optimal output.
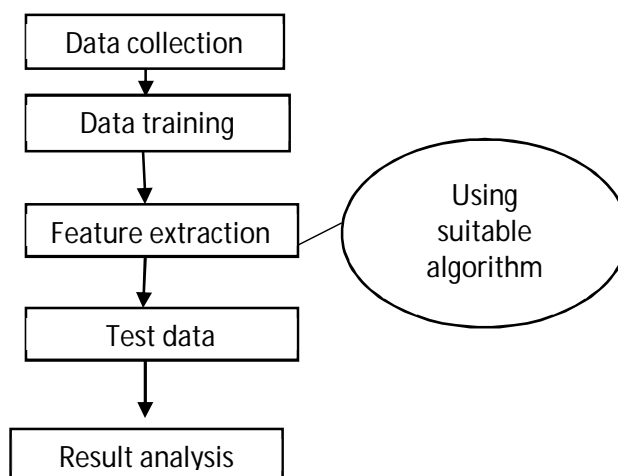

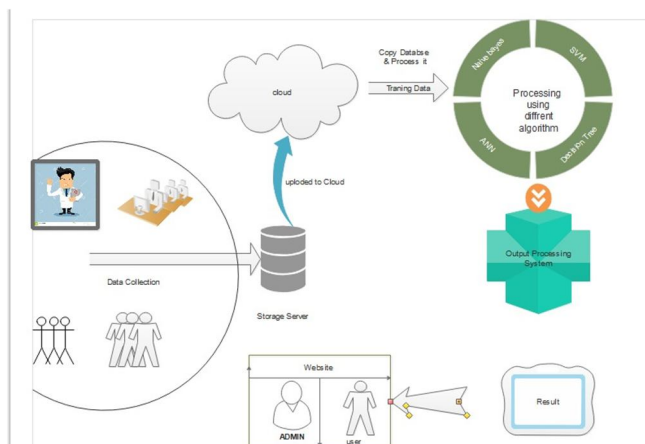Fig (1.1): Flowchart

IV.   ARCHITECTURE

Fig (1.2): The above architecture gives us a clear picture of proposed system.

V.   CONCLUSION

At last, we expect that we have reached our results of predicting diabetes by applying machine learning algorithms widely. The entire procedure was based on machine learning workflow. We have completed a classification model task with proper practical knowledge to deal with. We have undergone through many steps such as data exploration, data cleaning steps, feature of engineering basics and advanced model feature selections, model selection parameters and hyper parameter tuning using Scikit Learn library from Jupyter notebook. There is a strong possibility that running the project with logistic regression would give much better results.

REFERENCES

[1]   Priyanka Sonar, Prof. Jayamalini, "Diabetes prediction using different machine learning approaches", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.

[2]   Aishwarya Mujumdar, Dr. Vaidehi, "Daibetes Prediction using ML Algorithms", INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019.

[3]   Dr. Kayal Vizhi, Aman Dash, "Diabetes Prediction using Machine Learning", International Journal of Advanced Science and Technology Vol. 29, No. 6, (2020), pp. 2842 – 2852.

[4]   Quan Zou, Kaiyang Qu, Yamel Luo, Dehul Yin, Ying Ju, Hua Tang, " Predicting Diabetes Mellitus with machine learning techniques" November 2018 | Volume 9 | Article 515.

[5]   Bhavya, Sanjay, Suraj, Savant, Sanjay M, "Daibetes prediction using ML", ISSN (Online) 2278-1021 ISSN (Print) 2319-5940 International Journal of Advanced Research in Computer and Communication Engineering Vol. 9, Issue 7, July 2020.

[6]   Sakshi Gujral, "Early Diabetes detection using ML: A review", IJIRST –International Journal for Innovative Research in Science & Technology| Volume 3 | Issue 10 | March 2017 ISSN (online): 2349-6010.

[7]   Pramila Chawan, "Diabetes Prediction Using Machine Learning Techniques", Article · January 2018 DOI: 10.9790/9622-0801020913.

[8]   Naveen Kishore, Rajesh, Vamsi Akki, "Prediction Of Diabetes Using Machine Learning Classification Algorithms", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 01, JANUARY 2020 ISSN 2277-8616.

[9]   Hang Lai, Huaxiong, Karim, Aziz, Xin, "Predictive models for diabetes mellitus using machine learning techniques" Lai et al. BMC Endocrine Disorders (2019) 19:101 https://doi.org/10.1186/s12902-019-0436-6.

[10]   Amani, Akhtar, Jawad, Mirsat, " A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques", 978-1-7281-3992-0/19/$31.00 ©2019 IEEE.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ○ (24*7 Support on Whatsapp)