

Early Diabetes Prediction using Machine Learning

R S Geethanjali¹, Anusha G S², Sengeetha T D³, Sharath Kumar N⁴, Sumanth K Naik⁵

¹(Asst. Prof, CSE, KS School Of Engineering and Management, Mallasandra

Email: geethanjali.rs@kssem.edu.in)

² (CSE, KS School Of Engineering and Management, Mallasandra

Email: anushagowda710@gmail.com)

³ (CSE, KS School Of Engineering and Management, Mallasandra

Email:sangeethagowda2001@gmail.com)

⁴ (CSE, KS School Of Engineering and Management, Mallasandra

Email:sharath02001@gmail.com)

⁵ (CSE, KS School Of Engineering and Management, Mallasandra

Email: sumanthknaik@gmail.com)

Abstract:

Diabetes is a disease which lasts longer with the potential to cause a worldwide health care disaster. According to research there are 382 million people who are living with diabetes across the whole world. By 2035, this will be raised to 640 million. Diabetes is a disease caused due to the increased blood glucose level. This high blood glucose produces the symptoms such as polyuria, polydipsia and polyphagia. Diabetes is one of the leading causes of blindness, kidney failure, amputations, heart failure and stroke. When we consume food, our body converts food into glucose. At that point, our pancreas is supposed to produce insulin because it allows us to use the glucose for energy. The patients those suffering with diabetes, required insulin will not produce. Diabetes are the most common forms of the disease, but there are also different types, such as type 1, type 2 and gestational diabetes. Machine learning is an fast growing scientific field in data science which imitates human intelligence and learn from past experience. The aim of this project is to develop a model which can perform early prediction of diabetes for a patient with a greater accuracy by comparing the results of different machine learning techniques. The algorithms like K nearest neighbor, Logistic Regression, Random Forest and Artificial Neural Network are used. The accuracy of the system using each of the machine learning algorithms is evaluated. Then the one with a best accuracy is taken as the model for predicting the diabetes, The accuracy we expect from our project is above 95%.

Keywords —K nearest neighbour, Logistic Regression, Random Forest and Artificial Neural Network.

I. INTRODUCTION

Diabetes is a disease which lasts longer with the potential to cause a worldwide health care disaster. Many algorithms are used to predict diabetes, including machine learning methods like Random Forest, (KNN) K-Nearest Neighbor, Decision Tree and so on. With this machine

learning techniques. Diabetes mellitus or imply diabetes is a disease caused due to the increase level of blood glucose and by the parental genetics (DNA). Traditional methods of different kinds, based on physical and chemical tests, are available for diagnosing diabetes disease. Early prediction of diabetes is bit challenging task for medical field due to complex interdependence on different parameters as diabetes affects human

organs such as eye, heart, kidney, nerves and foot. Machine learning is a fast growing scientific field in data science which imitates human intelligence and learn from past experience. The aim of this project is to develop a model which can perform early prediction of diabetes for a patient with a greater accuracy by comparing the results of different machine learning techniques.

II. METHODOLOGY

In order to achieve our goal, our methodology comprises a few steps from which we accumulate datasets of the given attributes for the patients and we will do the pre-processing of our given attribute to apply on the given machine learning techniques find out the predictive analysis of the data.

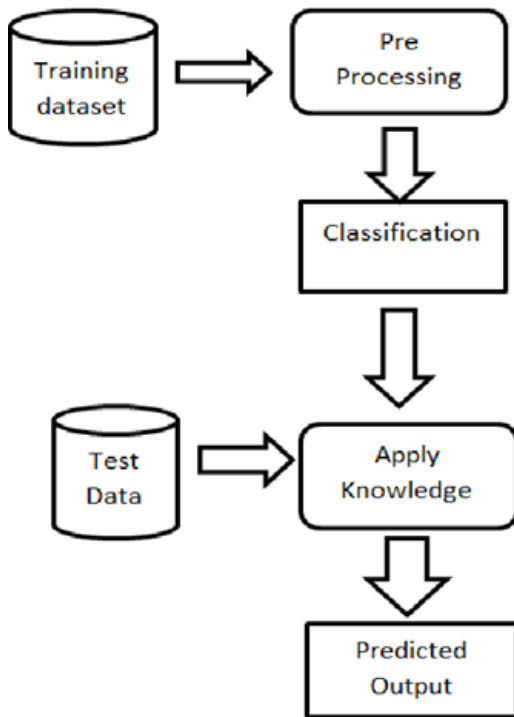


Fig 1: Disease Prediction Algorithm

A. DATASET AND ATTRIBUTES

In this work, we collect diabetes data from Kaggle. The dataset consists of various attributes for diabetes mellitus for 2000 patients. the attributes are.

S no.	Attributes
1	Pregnancy
2	Glucose
3	Blood Pressure
4	Skin Thickness
5	Insulin
6	BMI (Body Mass Index)
7	Diabetes Pedigree Function
8	Age

Table 1: Dataset Description

B. APPLYING MACHINE LEARNING TECHNIQUES

Once the data has been created for modelling we employ our four machine learning classification algorithm which we are going to implement to predict diabetes mellitus. Some overview of these techniques.

1. RANDOM FOREST:

Are an ensemble learning method for classification and regression and other task that operates by constructing a multitude of decision tree at training time and outputting the class that takes average votes of individual trees. The first algorithm for random decision forests was created by Tin Kam Ho using random subspace method. Ho established that to gain the accuracy it should over train where it can randomly restrict sensitive selected features of the given data.

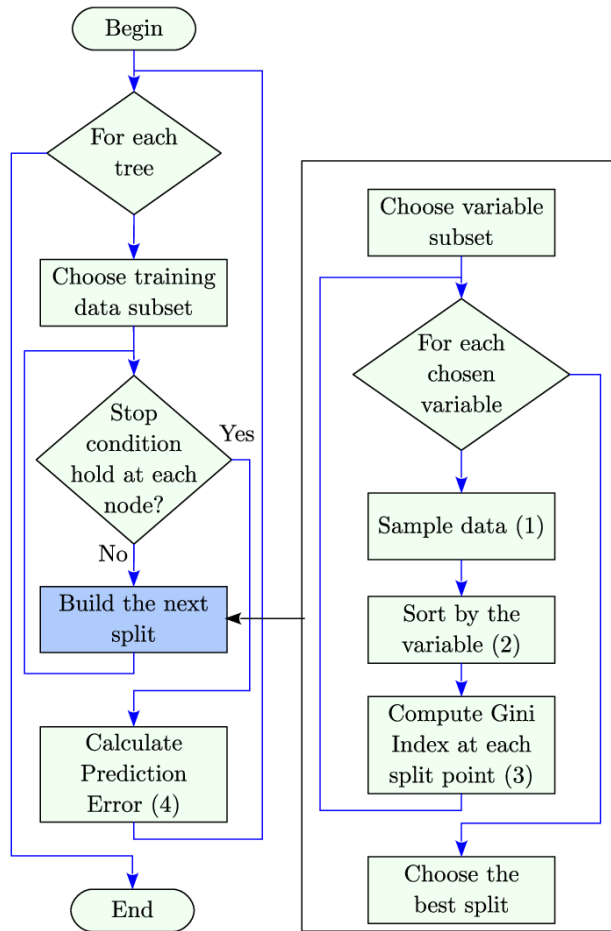


Fig 2: Working of Random Forest Algorithm

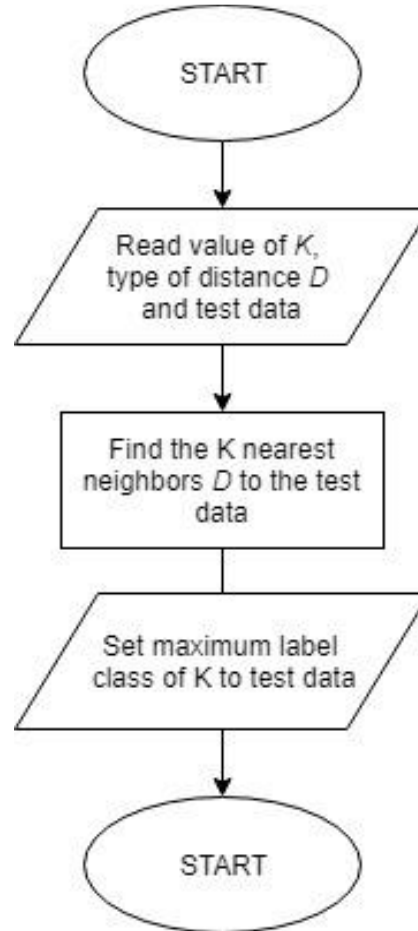


Fig 3: Working of K Nearest Neighbor Algorithm

2. K-NEAREST NEIGHBOR:

The classification and regression are used via k-nearest neighbor without parametric method. The input consists of k-closest training example in the feature space. To determine the distance from point of interest to point of training data set it uses. In classification technique, the value of k is always a positive integer of the nearest neighbor. The characteristic of the K-Nearest Neighbor algorithm is that it is sensitive to the local structure of the data.

3. Artificial Neural Network (ANN)

As ANN method is similar to the biological neural network of humans. It consists of neurons and connections with weights that are controlled to achieve proper output. An artificial neural network contains three layers. The input layer takes the inputs and their probability for computing the model. The hidden layer gives the weight to the input's probability by taking the output of the input layer. The output of the neural network will provide the neurons of the output layer.

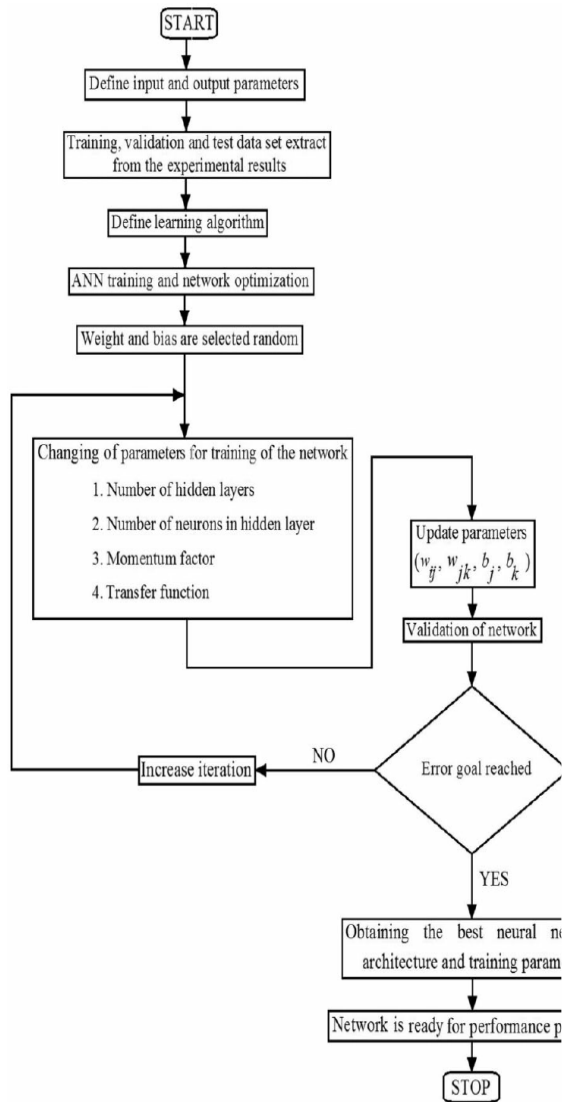


Fig 4: Working of Artificial Neural Network Algorithm

4. Logistic Regression

Logistic regression is a machine learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and it is based in the concept of probability. We can call a logistic regression a linear regression model but the logistic regression uses a more complex cost function, it is also known as logistic function. The cost function of hypothesis of logistic regression will range from 0 and 1. Therefore linear functions fail to show it, as it can make an outcome for the given function to

the possible hypothesis if it is greater than 1 and if it is less than it is 0.

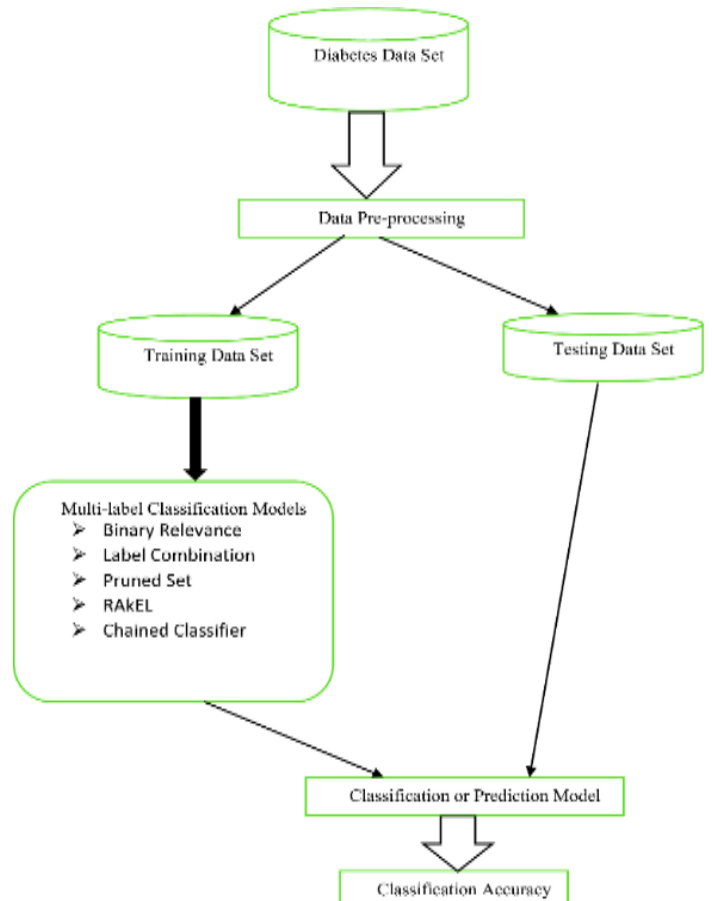


Fig 5: Working of Logistic Regression

The figure shows the overall process of the work which is being implemented. After defining the given problem we processed the data for constructing a prediction model. After applying the pre-processed data we get a training dataset and testing dataset. Now after applying the given machine learning techniques we get the test data which is being formed will now give us the performance of the techniques, thus by providing us with the best classifier of predicting diabetes mellitus.

GOALS AND OBJECTIVES

The goal of this project is to find effective machine learning based classifier models for detecting diabetes in individuals utilizing clinical data. The results of this study suggest that an appropriate pre-processing pipeline on clinical data and applying ML based classification may predict diabetes accurately and efficiently. The derived results for predicting the diabetes will be displayed using smart web application.

[1] To collect dataset of different classes. Indian diabetes dataset named PIMA will be used for analysis for this study. It consists of seven independent attributes. [2] To improve quality and effectiveness of obtained data. In order to increase the performance of machine learning models the accuracy must be improved. [3] To predict diabetes using ML algorithms. Machine Learning algorithms such as logistic regression, KNN, SVM, Random Forest classifier because researchers have proven demonstrated that machine learning algorithms are more effective. [4] To provide information to the patients whether they may affect diabetes or not. The technique may also help researchers to develop an accurate tool that will help them make better decision about the disease status.

CONTRIBUTION TO SOCIETY

The model could be used in health monitoring programs, and as an automated mass population screening tool without the need for extra costs compared to traditional methods. It increases the risk of early death, and its complications can lower quality of life. Early prediction of diabetes can help you avoid complications. Since diabetes is a non-curable disease, the present treatments and medications will only control associated symptoms. Therefore, it can significantly help diabetes research and, ultimately, improve the quality of healthcare for diabetic patients. It allows you to

understand the importance of diet plan and exercise along with the medications

APPLICATIONS

This application provides wide range of benefits for doctors as well as patients as it provides early prediction of diabetes, it provides easy accessibility for both the doctors and patients. This system also helps patients to check their results using webpage without consulting doctors. Diabetes prediction using machine learning is to provide warnings to patients who are at the starting stage of diabetes. It helps people to take precautions when they grow older. It also helps doctors to predict diabetes via machine learning algorithms using past medical records of diabetes patients. As we are developing the application, it is accessible on user's device such as smart phones and computers. It will provide result with few known parameters instead of blood test.

CONCLUSION

The main aim of this project is to present a clear overview of automatic Machine Learning based diabetes prediction. Various ML techniques in the prediction of diabetes were analysed in this review, which have been developed in recent days for the effective and efficient prediction of diabetes. The goal of developing a diabetes prediction model is to shift from higher precision to higher reliability for real time applications. The model is trained and tested with only few techniques. Since Logistic Regression, Random Forest and ANN is increasing worldwide, a model that can be used to predict diabetes in the world population is needed. Hybrid ML/combined algorithms can be introduced in the challenging complicated datasets in future research to obtain better performance. These techniques will be helping patients when performing real-world tasks in digital reality. Analysis of Diabetes disease is an attractive area for big data Analytics research for various purposes, including its significant impact on health and the demand for diabetes prediction. In the future, if more and more diabetic

patients use automatic glucose sensors that continuously measure glucose levels, the amount of data related to blood sugar will greatly increase. This discussion helps to provide a clear-cut view of diabetes prediction and helps to frame better diabetes prediction techniques to overcome diabetes through timely prediction.

ACKNOWLEDGMENT

We are doing this work under the mentorship of Dr. K Venkata Rao (Professor & Head) & Mrs. R S Geethanjali (Assistant Professor), Department of Computer Science & Engineering at K S School of Engineering and Management. We are doing a Project on Machine Learning where we have learnt the various Machine Learning Algorithms. We would like to express our special thanks to our mentor for inspiring us to take up the project & write this paper. Without their active guidance, help, cooperation & encouragement, we would not have my headway in writing this paper. We are extremely thankful for their valuable guidance and support on completion of this paper. We extend our gratitude to “K S School of Engineering and Management” for giving us this opportunity. We would also like to thank our parents for their support.

REFERENCES

- [1] A Review of Diabetic Prediction Using Machine Learning Techniques | M. Rajeswari 1, Dr. P. Prabhu 2 | (M.Phil research scholar, Department of Computer Applications, Alagappa University, 2 (Assistant professor in IT (DDE), Department of Computer Applications, Alagappa University, Karaikudi. | published on 2019.
- [2] Diabetes Prediction using Machine Learning Techniques | Mitushi Soni Dept of Computer Science and Engineering Shri G.S. Institute of Technology and Science, Dr. Sunita, Shri G.S | September-2020 | Published by IJERT.
- [3] DIABETES PREDICTION USING MACHINE LEARNING | Kishan Patel Department of IT Engineering PHCET, Manu Nair Department of IT Engineering PHCET, Shubham Phansekar Department of IT Engineering PHCET | International Journal of Scientific & Engineering Research | Volume 12 | Issue 3 | March-2021.
- [4] An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning | Umm e Laila 1, Khalid Mahboob 2, Abdul Wahid Khan 3, Faheem Khan 4, and Whangbo Taekun 4, Department of Computer Engineering, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan; | published at sensor | Published: 13 July 2022.
- [5] Diabetes Prediction Using Machine Learning | International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue IV Apr 2022 | Ashwini R1, S M Aiesha Afshin2, Kavya V3, Prof. Deepthi Raj.
- [6] Dr. Vaidehi V, Aishwarya Mujumdar “Diabetes Prediction Using Machine Learning algorithms” Year: 2019 | Journal Article | Publisher: ELSEVIER
- [7] Md. Kamrul Hasan, Md. Ashraful Alma, Dola Das, Eklas Hossain, Mahmudul Hasan “Diabetes Prediction Using Ensemble of Different Machine Learning Classifiers” IEEE Access Year: 2020 | Volume: 8, Journal Article | Publisher: IEEE
- [8] Jitranjan Sahoo, Manoranjan Dash, Abhilash Pati “Diabetes Prediction Using Machine Learning Classification Algorithms ” Year: 2020 | Volume: 7, Journal Article | Publisher: IRJET
- [9] Leila Ismail, Huned Materwala “intelligent diabetes mellitus prediction framework using machine learning” emerald Access year: 2021 | Journal Article | Publisher: EMERALD ISSN: 2634-1964
- [10] Usama Ahmed, Ghassan Issa, Muhammad adnan Khan, Shabib Aftab, Muhammad Farhan Khan, Reaed A T Said, Taher M Ghazal “Diabetes Prediction Empowered with Fused Machine Learning” | Year: 2022 | Volume :5, Journal Article | Publish: IEEE
- [11] NIKOS FAZAKIS, OTILIA KOCSIS, ELIAS DRITSAS, SOTIRIS ALEXIOU, NIKOS FAKOTAKIS AND KONSTANTINOS MOUSTAKAS “Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction” | Year: 2021 | Volume: 8, Journal Article | Publisher: IEEE