IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 8, June 2022

Impact Factor: 6.252

Diabetics Prediction System UsingMachine Learning

Darshan Vaijanath Khatal[1], Ganseh Dinkarrao Kulkarni[2], Shraddha Satish Kshatriya[3],Prasad Bharat Jadhav[4]

NBN Sinhgad School of Engineering, Pune, Maharashtra, India[1,2,3,4]

Abstract: *Diabetes is a serious disease that can strike at any time and affect many people. Age, obesity, sedentary lifestyle, poor diet, and high blood pressure are just a few of factors that contribute to the development of type 2 diabetes. There is the number of health problems that are more common in people with diabetes than usual population. Patients with diabetes are currently being diagnosed and treated using a variety of diagnostic methods, including blood testing, urine tests, and more. In the healthcare industry, big data analytics is essential. The healthcare industry has a tremendous amount of value data stored on a website. By using large amounts of data, users can gain understanding and practice prediction ns about the future by exploring big data sets and hidden disclosures knowledge and styles. The current method is not very good at classifying and forecasting. To better classify diabetes, we present a diabetes prediction model in this article that incorporates a few extrinsic parameters that cause diabetes, as well as regular components such as glucose, creatinine ratio, urea, fasting lipid profile, body mass index, age, insulin, and so on. Both datasets, each with eight variables, were subjected to the identical tests. The accuracy of a dataset with 12 variables is higher, so the conclusion is that the more information we have, the more accuracy we can attain.*

Keywords: Diabetes, Machine Learning, Prediction, Dataset, etc.

I. INTRODUCTION

It is generally agreed that diabetes is the deadliest and most persistent condition that contributes to an increase in glucose levels. According to the International Polygenic Disease Federation, polygenic disease is the type of diabetes in which the exocrine gland is unable to produce hypoglycemic agent. There are 382 million people around the world who are affected by polygenic illness. That number will more than quadruple to 592 million by 2035.

Mellitus, also known as illness, is a condition caused by an increase in blood glucose levels. If diabetes is not treated andthe patient does not go to the doctor, a number of complications are likely to develop. The risks include harm to the excretory organs, which often results in a chemical analysis; damage to the eyes, which may result in vision impairment; or an increased risk of cardiopathy or stroke in associate degree patients.

After going through a rough patch, the diagnostic process, the patient will eventually have to go to a diagnostic center and talk to their doctor about getting further treatment. The growing field of machine learning offers a solution to this important challenge.

The use of the Random Forest algorithm as a form of machine learning technology, the purpose of this work is to build a competent system to make early diagnosis of diabetes in a patient with a high degree of accuracy using a specified algorithm in addition to being a form of ensemble learning method, random forest algorithms are frequently employed for each classification and regression problem. In addition, random forest algorithms are used. When compared to other algorithms, the level of accuracy is significantly higher.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

II. RELATED WORK

Diabetes is a chronic disease or group of metabolic diseases that causes a person to have an elevated level of blood glucose in the body for an extended period of time. This can occur because the body does not produce enough insulin or because the cells in the body do not respond to insulin in the appropriate manner. High blood sugar levels associated with diabetes are associated with chronic damage, deterioration, and failure of many organs, especially the eyes, kidneys, nerves, heart and arteries. Utilizing relevant features, developing a prediction algorithm by utilising machine learning, and determining the best classifier to utilise in order to achieve the most accurate results possible when compared to clinical outcomes are the characteristics that are useful in the early diagnosis of diabetes mellitus by making use of predictive analysis.

According to the findings, the decision tree algorithm and the Random Forest model are the ones that perform the best when used to the analysis of diabetic data. The respective specificities of these two models are 98.20 and 98.00 percent. According to the result of a naive Bayesian analysis, the level of accuracy achieved is 82.30 percent. Additionally, this research generalizes the selection of appropriate features from datasets in order to increase classification accuracy [1].

Along with common factors such as glucose, body mass index (BMI), age, and insulin, the authors the paper proposes a diabetes forecast model to better plan for diabetics. This the model considers a few external factors that cause diabetes in in addition to the usual features. Separation accuracy is improved by using new ones Database compared to an existing dataset. In addition to this, we have implemented a pipeline model for diabetes prediction with the goal of improving the overall accuracy of categorization [2].

One of the most dangerous and persistent diseases, diabetes is responsible for a rise in the amount of glucose (sugar) in the blood. If diabetes is not treated and the condition is not detected, it can lead to a number of issues. Because the procedure of identification is so laborious, it always ends with the patient going to a diagnostic centre and speaking with a physician. However, the development of methods based on machine learning has provided a solution to this crucial issue. The purpose of this investigation is to devise a model that can predict the likelihood of diabetes in patients with the highest possible degree of precision. In order to detect diabetes in its earliest stages, this experiment makes use of three different machine learning classification algorithms: decision trees, support vector machines, and naive bayes.

The test was performed on the Pima Indians Diabetes Database (PIDD), from a machine storage facility at the University of California, Irvine. The performance of each of the three algorithms is tested using a variety of metrics, including accuracy, precision, F measurement, and memory, among others. The number of positive and negative values is used to measure accuracy. The collected results show that Naive Bayes surpasses other algorithms, with a high accuracy of 76.30 percent achieved by it. These results are confirmed by the use of the Receiver Operating Characteristic (ROC) curves in a precise and practical manner [3]

The field of machine learning has had a considerable influence on a variety of subfields within science and technology, including the fields of life sciences and medical research. Diabetes mellitus, more frequently referred to as diabetes, is a chronic condition that is characterized by abnormally high levels of glucose sugar in blood cells and the utilisation of insulin in the human body. Diabetes mellitus is often referred to as diabetes. This article focuses on screening patients with diabetes and diagnosing diabetes using a variety of Mechanical Learning methods to create a low-resolution model based on the PIMA database.

These techniques are used to model. Both the experimental piece of PIMA and the data collected at Kurmitola General Hospital in Dhaka, Bangladesh, used the model validation process. This study is being carried out in order to demonstrate the efficacy of multiple classifiers that have been trained on the diabetes dataset of a specific nation and then evaluated on patients from patients from another country. During the course of this investigation, we tested various classification strategies, including decision trees, K- nearest neighbour analysis, random forests, and Naive Bayes.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

The findings indicate that random forests and Naive Bayes classifiers did well on all datasets [4]. Cardiovascular diseases, sometimes known as CVDs, are among the most common life -threatening illnesses that affect human health. Early detection can help prevent or treat cardiovascular disease (CVDs), which can lead to lower mortality. The application of machine learning models to the task of determining riskvariables is a method that shows promise. The authors would like to suggest a model that takes into account a variety of approaches in order to obtain accurate prediction of cardiovascular disease. In order for this model to be effective, we have used approaches that are efficient in terms of Data Collection, Data Preprocessing, and Data Transformation in order to provide reliable information for the training model [5].

III. PROPOSED SYSTEM

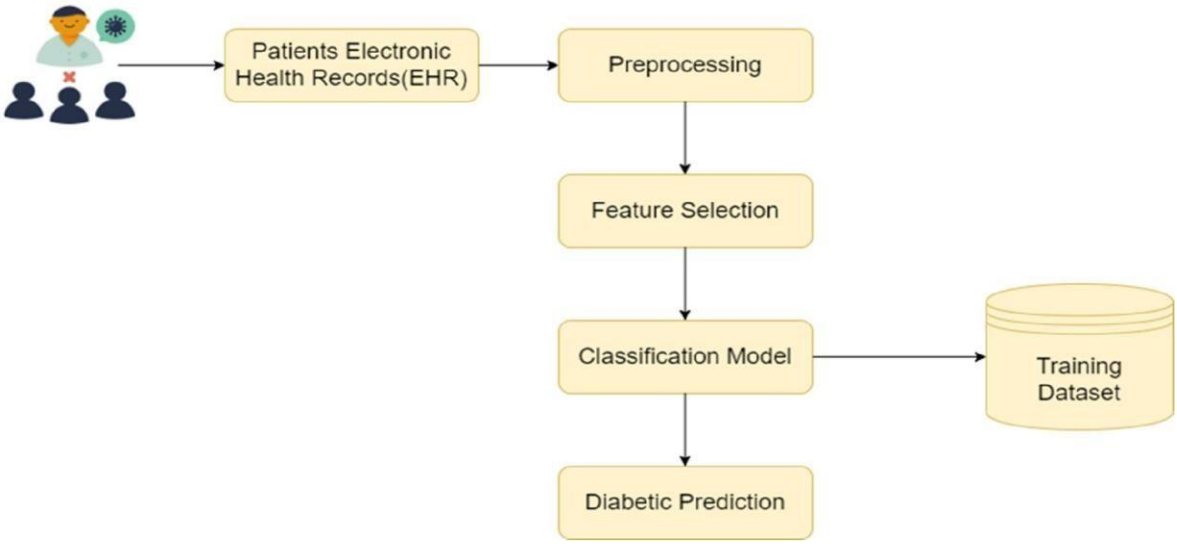

Figure 1: System Architecture

IV. CONCLUSION

- Using this method, you may determine the potential dangers that diabetes patients face by calculating their risk of having a stroke or a heart attack.
- It can be anticipated by making use of the diabetes risk factors that are prevalent in the population.
- The prediction of the attack of the disease is heavily dependent on the quantification of risks given by each element.
- It is vital to measure the risk variables in order to avoid the patients from being exposed to the dangers of the consequences.

REFERENCES

[1] N. Sneha and T. Gangil "Analysis of diabetes mellitus for early prediction using optimal features selection" Journal of Big Data vol. 6 no. 1 2019.
[2] Mujumdar and V. Vaidehi "Diabetes Prediction using Machine Learning Algorithms" Procedia Computer Science vol. 165 pp. 292-299 2019.
[3] D. Sisodia and D. Sisodia "Prediction of Diabetes using Classification Algorithms" Procedia Computer Science vol. 132 pp. 1578-1585 2018.
[4] B. Pranto S. Mehnaz E. Mahid I. Sadman A. Rahman and S. Momen "Evaluating Machine Learning

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

Methods for Predicting Diabetes among Female Patients in Bangladesh" Information vol. 11 no. 8 pp. 374 2020.

[5] P. Ghosh et al. Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques IEEE Access vol. 9 pp. 19304-19326 2021.