IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Dietetics Prediction System Using Machine Learning

Darshan Vaijanath Khatal, Ganseh Dinkarrao Kulkarni, Shraddha Satish Kshatriya, Prasad Bharat Jadhav
NBN Sinhgad School of Engineering, Pune, Maharashtra, India

Abstract: *Diabetes is a severe disease that can strike at any time and affect a large number of people. Age, obesity, sedentary lifestyle, poor diet, and high blood pressure are just few of the factors that contribute to the development of type 2 diabetes. There are a number of health problems that are more common among diabetics than in the general 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 colossal amount of data stored in databases. Using big data analytics, users can acquire insight and make predictions about the future by examining large datasets and uncovering hidden information and trends. The current method isn't 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

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 hypoglycaemic agent. There are 382 million people around the world who are affected by polygenic illness. This number will more than quadruple to 592 million by the year 2035. Diabetes mellitus, also known as simply illness, is a condition that can be caused by an increase in the glucose level of the blood. If the diabetes is not treated and the patient does not go to the doctor, there is a good chance that several complications will arise. 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 the laborious process of identification, a patient will ultimately have to go to a diagnostic centre and speak with their physician about receiving more treatment. The growing field of machine learning offers a solution to this fundamental challenge. Making use of the Random Forest algorithm as a type of the machine learning technique, the purpose of this work is to construct a system that is capable of performing early prediction of diabetes for a patient with a greater level of accuracy utilising said 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.

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. The persistently high blood glucose levels that are associated with diabetes are linked to the long-term damage, breakdown, and failure of a number of organs, most notably the eyes, kidneys, nerves, heart, and veins. 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 goals of this research. The objective of the suggested strategy is to zero in on the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

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 generalises the selection of appropriate features from datasets in order to increase classification accuracy [1].

Along with regular factors such as glucose, body mass index (BMI), age, and insulin, the authors of this paper propose a diabetes prediction model in order to better classify diabetics. This model takes into account a few external factors that are responsible for diabetes in addition to regular factors. The accuracy of classification is improved by using a new dataset in comparison 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. Experiments are run on the Pima Indians Diabetes Database (PIDD), which comes from the machine learning repository at the University of California, Irvine. The effectiveness of each of the three algorithms is assessed using a variety of metrics, including precision, accuracy, F-measure, and recall, amongst others. The number of instances that are correctly and wrongly categorised is used to measure accuracy. The findings that were collected indicate that Naive Bayes surpasses other algorithms, with the maximum accuracy of 76.30 percent being achieved by it. These results are validated through the application of Receiver Operating Characteristic (ROC) curves in an appropriate and methodical 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 characterised by abnormally high levels of glucose sugar in blood cells and the utilisation of insulin in the human body. Diabetes Mellitus is more generally referred to as diabetes. This article has concentrated on evaluating diabetes patients as well as the diagnosis of diabetes using a variety of Machine Learning techniques to construct a model with a few dependencies based on the PIMA dataset. These strategies were utilised to build up the model. Both an unexplored piece of PIMA and a dataset gathered from Kurmitola General Hospital in Dhaka, Bangladesh, have been utilised in the validation process for the model. 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. The findings indicate that random forests and Naive Bayes classifiers did particularly well on both datasets [4].

Cardiovascular disorders, sometimes known as CVDs, are among the most prevalent forms of life-threatening sickness that influence human health. Early detection may help in the prevention or treatment of cardiovascular diseases (CVDs), which in turn may lead to lower mortality rates. The application of machine learning models to the task of determining risk variables 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].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

III. PROPOSED SYSTEM

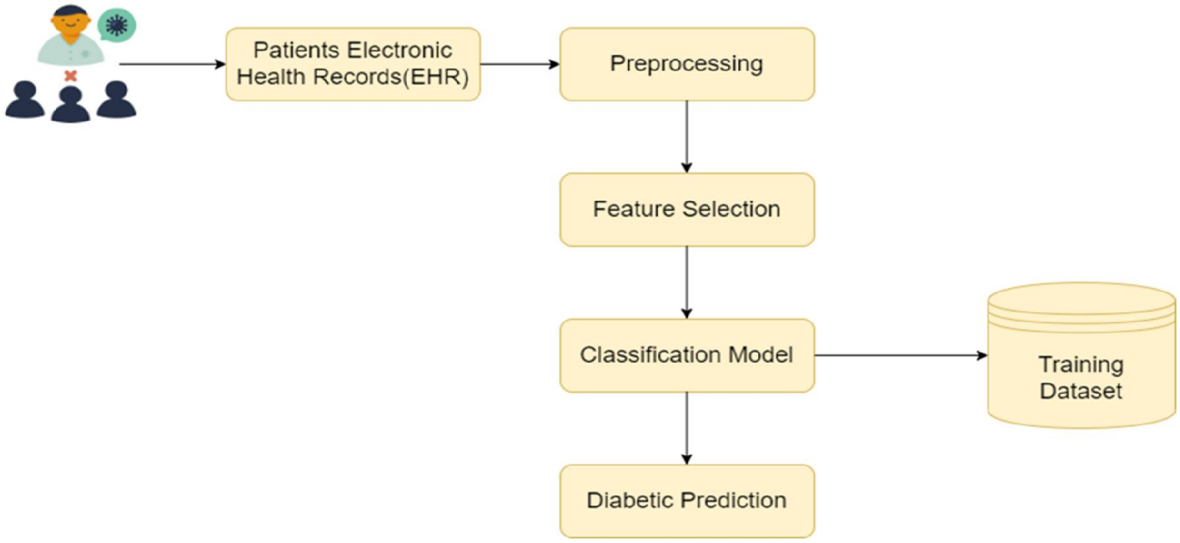

Figure 1: System architecture

III. 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 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