International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

DIABETES RISK ASSESSMENT USING MACHINE LEARNING

Anjali A P
PG, Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
anjalimca2025@gmail.com

Sowmya J
Assistant Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
sowmyaj@theoxford.edu

ABSTRACT

In this research paper represent an inclusive study on Diabetes risk assessment using machine learning which is one of the increasing global health need that needs timely assessment precluding care. Prior prediction and also the management are very important in the factor of reducing complication and also to improve the patient outcomes. This paper is going to represent a Diabetes risk assessment using Machine learning that leverages machine learning algorithm which is going to predict the likelihood of the diabetes by considering the factors which are based on patient health indicators such as the glucose levels, blood pressure, body mass index (BMI), insulin, and age. This system is going to combine the user authentication, and also provide the personalized diet planning, nearby clinic locator and an AI chat assistant to provide the comprehensive health support. This enables users to recognize suitable medical care for further evaluation of their health.

KEYWORDS: *Diabetes risk assessment, Hypertension, Blood glucose level, BMI, Early detection.*

INTRODUCTION

Diabetes mellitus commonly referred to as Diabetes is a metabolic condition that is diagnosed using blood sugar level. Going by the survey there are more than 537 million adults worldwide currently living with the diabetes. This insatiable diabetes may turn out to be the issue to cause serious complications like the cardiovascular disease, blindness and kidney failure. The project is exploited in order to establish the connection between patients and preventative care through the identification of personal health data with the help of machine learning models to provide the real-time possibility of diabetes risk and offer the diet plan There might also circumstance that

the patient will have the permission to access to the blood test report but they will not have the permission to access the results whether they are having positive or negative which also means that the particular person is has diabetes or not results in the report in such cases the artificial intelligence in collaboration with health care information is going to encounter the requirement as an increment, the predictive system which is built is going to give the prior warnings. The main scope in this undertaking is that to come up with a powered diabetes risk assessment based smart system which is going to predict the inclination of the diabetes also to have the ability to provide actionable information.

LITERATURE SURVEY

Lending money to the machine learning methods to the healthcare department has been a good idea particularly when it comes to predicting whether an individual has diabetes or not. Other people have yet gone a step further to perfect the quality of such predictions even more in various manner. The other model of the machine known as the Logistic Regression machine learning algorithm may have been used quite generically and in cases where he was used generally then the machine may not have recorded appropriate expectations. It has been observed that SVM works on a high volume of data set, though DT and RFM also achieved fair results as they are not tuned fast and do not capture the trends in bending or turning.

Strengths and weaknesses were on the use as steered by Santhanam and Padmavathi (2015) who employed the use of clustering and decision tree tools and produced effective results, but with weaknesses on the large scale of application. In 2017, Kavakiotis and other authors have conducted a review during which the authors stated that to perform the analysis of the issue of the threat of diabetes, Random Forest, in addition to group approaches, may be a reasonable solution to use. The inner-type learning observed in the large brain network could also be exploited and they could be more precise although much information may be involved as well as the computer backbone. The majority of the methods available today are only concerned with the prediction but do not offer any other kind help like some of the ways and means to live well or/and the means to receive care without inconvenience. When comparing the solution that we have developed to the mentioned aspects, our proposal is an extension of the above because it is an integrated solution that includes machine

learning forecasting, personal dietary program, hospital directory and AI support.

EXISTING WORK

Machine learning has predicted diabetes successfully by a variety of researchers to varying degrees. Firstly, because it is easy and understandably was Logistic Regression. There has been some effort, also with SVM and k-nearest neighbors (KNN). Their performance will be adequate, good enough but they will not perform on the noisy data or the unstructured data other improved better models can be found such as the Decision Tree for instance Random Forest, it is never going to give overspecialized and handle large volume of complicated data. Using the cases of Pima Indian Diabetes dataset, we can assume or analyze that success rate is somewhere in between 70 and 85 percent and it is going to depend on how it is being prepared and completed. It is being considered with the subdivision of the information and deletion of client focused service oriented, involving the offer of healthy lifestyle, the current going screening or the cooperation with the health system. Despite the fact that the results of the current study have identified the possibility of AI to predict diabetes, two dimensions.

PROPOSED SYSTEM

The Endorsed system is the Diabetes Risk Assessment System that will function as a extensive package that is going to guide the identification of a risk in diabetes and in the mentorship of the patient and the stipulation of health to a patient. It is accomplished in Random Forest. It is alerted about the Pima Indian Diabetes. This scrutinizes like the level of sugar in the blood, the body mass index, insulin, blood pressure, whether the person was pregnant and age. The accessibility of a login and sign-up module make it simple in the overall use of system and also hypothesizing the safe character in a convoluted way. Besides guessing, the health status of the users guides dietary guidance adopted by the system and it advises diabetic and non-diabetic users. Moreover, the health services will become accessible due to the clinic location and Folium map.

Constant inspection and personal contact are also implemented in the system. These types of records could be observed over a period. The introduction of a visual board will enable the potential visualization of the presented data on charts and graphs to the patients through the chatbot.

METHODOLOGY

The way how we establish the architecture of our system Diabetes Risk Test System begins by choosing data to be utilized. We selected the Pima Indian Diabetes data that contains important health information such as sugar level, index of body fat, insulin, blood pressure, and age. We did replace incomplete information, levelled the inputs and divided data into learning and testing parts.

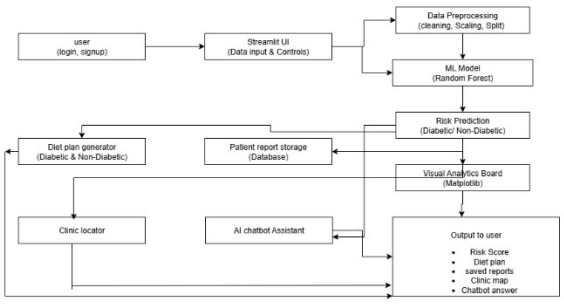

Fig. 1. Block diagram.

In this algorithm, to make guesses we selected Random Forest type of the classifier to estimate because it is able to model complicated relationships in the data and avoids intensive fitting. We have created a system with Streamlit on the user side, and then giving the user the opportunity to enter values and even get the result instantly. We also added additional functionality, i.e. generation of diet plans, storing reports in a database. into charts. This is the sense that such a mode of operation makes the system befitting not just in what is wrong and in calculating the unfitness's, but also good and fruitful whenever it happens to cope with diabetes in the long-term. Help users see how their health changes over time. This way of doing things makes sure the system is not just right in finding out what is wrong.

Task	Task Name	Status
1	Collect Diabetes Dataset	Done
2	Train Machine Learning Model	Done
3	Data Preprocessing & Cleaning	Done
4	Develop Prediction Module	Done
5	Implement Diabetes prediction System	Done

EXPERIMENTAL RESULTS

In order to test we selected the Random Forest tool of guessing since it is robust and powerful with complex data relationships. During the training period, we drove the tool by modifying parameters such as number of counters, and max depth. We were correct in our last tool 81.5% of the time, and outperformed other simple tools including Logistic Regression (77.8%) and Decision Tree (74.6%). We also tested the precision, recall and F-score in order to get a closer check. The precision, recall 0.82, and F1-score of the tool were 0.79, 0.80, respectively. This indicates that it is effective in right finding diabetic and non-diabetic. The additional components of the system were tested not only in guessing right. The diet plan creator has provided good food advice on diabetic and non-diabetic individuals. The clinic finder placed diabetes clinics in Bangalore on the map (right), and the AI chatbot provided decent responses to frequently asked questions of how to stop diabetes and what is the best way to live well. In addition, the user tests revealed that the inclusion of new features made it more fun to use and more practical and is promising when it comes to the entire online health repair.

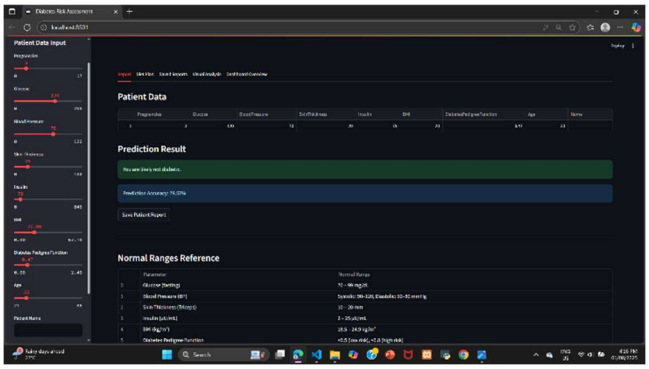

Fig. 2. Output of the diabetes prediction page

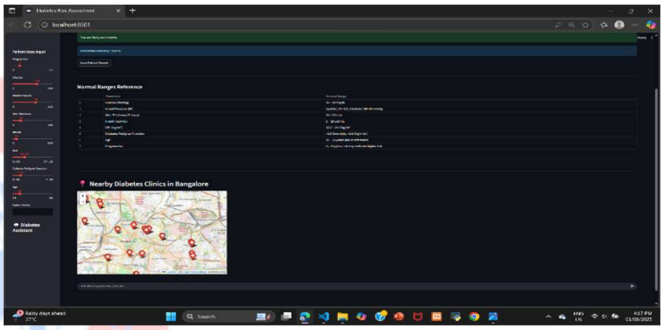

Fig. 3. It gives the map and chatbot for guidance

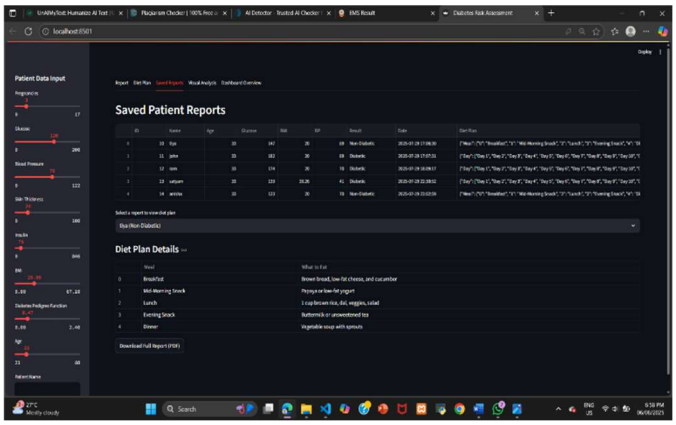

Fig. 4. Multiple saved reports which can be viewed.

CONCLUSION

Throughout this study, our goal was to provide a complete Diabetes Risk Test Tool was developed as an end-to-end kit that combines machine learning guess with care tools to give people something to handle how diabetes is running the hard round. Employing information in the Pima Indian Diabetes set, the tool applied a Random Forest pick way to predict who could have diabetes based on central body realities of sugar, BMI, blood push, insulin, and age. Exams demonstrated that this path was truer than old ways proving that it is superior in early detection of risk of diabetes.

In conclusion, the intended tool does not only present a fool-proof guessing method but a complete health heals digitally. Putting intelligent technology together with actual health requirements, it sits in the gap between find and care within human beings; therefore, a large and robust initiative to deal with a risk of diabetes. The tool put in many help bits like a made-for-you diet plan maker, a place to keep patient files, a look-at data board, a map to find clinics, and an AI talk tool. Tests by users showed more about how well and useful the tool is, noting how it can help both people and health work folks.

REFERENCES

[1] "Pima Indian Diabetes Dataset," Kaggle.https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database. [Accessed: Aug. 12, 2025].

[2] T. Santhanam and M. S. Padmavathi, "Use of K-means and genes codes for short size by using SVM for diabetes checks," Procedia Computer Science.

[3] D. Sisodia and D. S. Sisodia, "Classification algorithms in the prediction of diabetes," Procedia Computer Science, vol. 132, pp. 1578–1585, 2018. doi: 10.1016/j.procs.2018.05.225.

[4] D. S. Choubey and S. Paul, "Diabetes prediction using machine learning algorithms," Materials Today: Proceedings, vol. 33, pp. 4902–4906, 2020. doi: 10.1016/j.matpr.2020.04.466.

[5] F. Miao, X. Wen, Y. Li, and Y. Li, "Machine learning for diabetes detection: A review," Health Information Science and Systems, vol. 8, no. 1, pp. 1–9, 2020. doi: 10.1007/s13755-020-00104-8.