World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

Diabetes mellitus detection using machine learning techniques

Samridhi Puri [1, 3, *], Satinder Kaur [1, 3], Satveer Kour [1, 3] and Kumari Sarita [2, 3]

[1] *Department of Computer Engineering and Technology, India.*
[2] *Department of Computer Science, India.*
[3] *Guru Nanak Dev University, Amritsar, India.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 12(01), 059–064

Publication history: Received on 28 March 2024; revised on 05 May 2024; accepted on 08 May 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.12.1.0181

Abstract

Diabetes Mellitus (DM) is a common disease that is spreading worldwide and affecting millions of people. The use of machine learning techniques for the timely and accurate diagnosis of diabetes is the subject of this study. This research makes use of a dataset that includes measures like insulin resistance metrics and blood glucose levels in addition to clinical and demographic data. The important factor leading to the growth of diabetes mellitus is the lifestyle which includes lack of exercise, anxiety, age factor, family history etc. It is a serious condition, if not cured, can lead to several health problems like heart disease, nerve damage, kidney damage. A range of machine learning algorithms and deep learning techniques such as support vector machines, decision trees, and neural networks, are utilized to examine and simulate the connections present in the data. Techniques for feature extraction and selection are applied to improve the models' performance and highlight important factors that lead to the development of diabetes. The goal of the study is to attain high sensitivity and specificity to guarantee the accurate identification of Type 1 and Type 2 diabetes.

Keywords: Diabetes Mellitus; Machine Learning; SVM; KNN

1. Introduction

As stated by the World Health Organization (WHO), around 422 million people worldwide have diabetes, which is expected to grow by about 578 million by 2030. (Gourisaria et al. 2022) Therefore, it is very important to predict diabetes mellitus. The National Diabetes Statistics Report 2020 USA, states that every 10 diabetic cases in United States go undiagnosed while other report from WHO confirmed that every 11th person on the planet is diabetic. Basically, diabetes is divided into two parts Type1 (T1DM) and Type2 (T2DM) diabetes mellitus. In Type1 diabetes, the pancreas does not produce insulin and in Type2 diabetes, the pancreas produce less insulin and body becomes resistant to insulin. T1DM is classified as a chronic autoimmune disease that is the result of elevated blood sugar levels (hyper-glycemia) where glucose level > 180mg/Dl (Afsaneh et al. 2022). The most common type is T2DM as it can be seen in 90% of cases. The early diagnosis and treatment of type 2 diabetes are among the most relevant actions to prevent further complications like diabetic retinopathy. Prediabetes can be determined by one of the following conditions: elevated glycated Haemoglobin A1c(HbA1c) levels, and Impaired Glucose Tolerance (IGT). T2DM is heritable and probability of getting T2DM is higher in siblings of a T2DM patient than in families without any T2DM patient(Afsaneh et al. 2022) Also, the risk of T2DM is elevated with a body mass index (BMI) >30 mg/Dl.

Early detection of diabetes mellitus can lower the risk of cardiovascular events, according to the ADDITION-Europe Simulation Model Study (Sharma and Shah 2021). A person is classified as having diabetes if their blood glucose level is higher than 70–100 mg/dl, prediabetes if it is between 100–125 mg/dl, and normal if it is between these ranges. Diabetes can cause several symptoms, such as increased thirst, increased hunger, weight loss, and blurred vision and if not cured can lead to several other problems.

* Corresponding author: Samridhi Puri

In today's era, it becomes easy to detect diabetes mellitus with the help of machine learning. Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computer systems to improve their performance on a specific task through learning from data. The use of various computer algorithms that can be improved on their own through testing and data utilization is known as machine learning. According to the algorithms produce a model that makes decisions based on sample data, or training data. Numerous studies have been conducted utilizing various machines.

Various researchers have discussed numerous classification models such as logistic regression (LR), decision tree (DT), Random Forests (RF), support vector machine (SVM), and K-nearest neighbours (KNN) (Sharma and Shah 2021). Using ML techniques, the best detection approach for diabetes mellitus can be found (Fregoso-Aparicio et al. 2021) .

The number of studies developed in this field created two main challenges for the researchers is to build predictive model for type 2 diabetes (Firdous, Wagai, and Sharma 2022). The algorithms used for machine learning or any other field of artificial intelligence perform predictive modelling based on historical data (Rajeswari and Vijayakumarponnusamy 2021). Due to the complexity and variability of diabetes detection, a systematic framework for decision-making has been established and consists of steps. (Tech and Rathore, n.d.)

1.1. Datasets

Different datasets of patients are available on sites like Kaggle where various Machine learning algorithms can be applied to check the accuracy of the algorithms.

1.2. Preprocessing

The datasets are subjected to several pre-processing techniques before being fed into the machine learning model, ensuring that the model's performance is enhanced. Pre-processing entails eliminating outliers and handling encoding, data standards, missing values, and other issues

1.3. Feature extraction

The features are selected and then machine learning techniques are applied on it .

1.4. Use of Machine learning in detection of Diabetes

How different machine learning techniques can be used for diabetes mellitus detection.

Different Machine learning algorithms like SVM, KNN, Random Forest, Decision Tree are applied on the datasets.

1.5. Performance parameters

The algorithm that gives the highest accuracy is considered best for the diabetes mellitus detection.

The paper is organised in this order Section 2 delivers Literature review of the research papers, Section 3 provides proposed work, Section 4 provides methodology, Section 5 discusses results and analysis and Section 6 covers conclusion.

2. Literature review

Birjais et al. (5) experimented on PIMA Indian Diabetes (PID) data set which has 768 instances and 8 attributes and is available in the UCI machine learning repository. They aimed to focus more on diabetes diagnosis, which according to World Health Organization (WHO) in 2014, is one of the world's fastest growing chronic diseases. Gradient Boosting, Logistic Regression, and Naïve Bayes classifiers were used to predict whether a person is diabetic or not, with gradient boosting having accuracy of 86%, logistic regression having 79% accuracy, and naïve bayes having 77% accuracy. (5). In most of the research works, Pima Indian Diabetes Dataset (PIDD) have been used by many researchers for diabetes detection. (11)

Significant research has been done by various scientist for early detection of diabetes mellitus. The Machine learning approach is critical for predicting several medical databases, including hypertension data. (3) As concluded from various previous researches the main goal is how to check which Machine learning technique is the best for diabetes mellitus detection at an early stage. Previous researches show that Decision Tree model is the most suitable model in terms of accuracy. Various super-vised machine learning algorithms were used to predict diabetes. (11)

2.1. Proposed work

There is a discussion of the function of an algorithm and classification ensemble machine learning approach. The J48 decision tree was utilized to determine hypertension in patients with or without diabetes based on diabetes risk factors. The results of the study demonstrate that the Machine Learning ensemble technique is more efficient than bagging and a J48 decision tree. (3)

Islam et al. employed a variety of algorithms, including RF and LR, to analyse a dataset of diabetes symptoms. The dataset is first inserted into the system, where predictive models are then used, and accuracy is then noted. Using algorithms, Choudary et al. divided the population into two groups: high-risk individuals and low-risk individuals. They employed methods like DT, RF, and soon after to cluster data into groups and utilized SVM to create a hyperplane. Shukla et al. predicted the accuracy using the LR algorithm and a dataset (Rajeswari and Vijayakumarponnusamy 2021). PID, Case 1 and PID, Biswas, and Bandyopadhyay 2022 were the datasets used by Dalakleidi et al. The two datasets used by Dalakleidi et al. (Kaur and Kumari 2022) were PID, Case 1, and Hippokrateoin, Case 2 from so different algorithms, including as Decision Tree, SVM, and Random Forest, have so far proven to be the most effective models for early diabetes identification. The main contribution of this article is its consideration for ML based methods for DM detection, diagnosis, self-management, and personalisation.

3. Methodology

The methodology gives an overview of the different platforms used by researchers like Google Scholar, Scopus, Web of Science and Research Gate to collect and extract data with the help of keywords like Diabetes Mellitus, Machine learning. The data is searched from 2018 to 2022. A total of 40 papers were collected 20 from Google Scholar, 10 from Web of Science and 10 from Research Gate out of which only 20 papers were extracted for the review 10 from Google scholar, 5 from Web of Science and 5 from Research Gate. The main questions to be considered are:

- **RQ1** How Machine learning can be used to detect Diabetes Mellitus at an early stage?
- **RQ2** Which Machine learning model is best for diabetes detection?
- **RQ3** What are the optimal validation metrics to compare a model's accuracy?

3.1. Experimentation

Basically, visualization of dataset is done for diabetes prediction using python and dataset is collected from PIMA Indian Diabetes (PID) data set which has 768 instances and 8 attributes.

3.2. Results and analyis

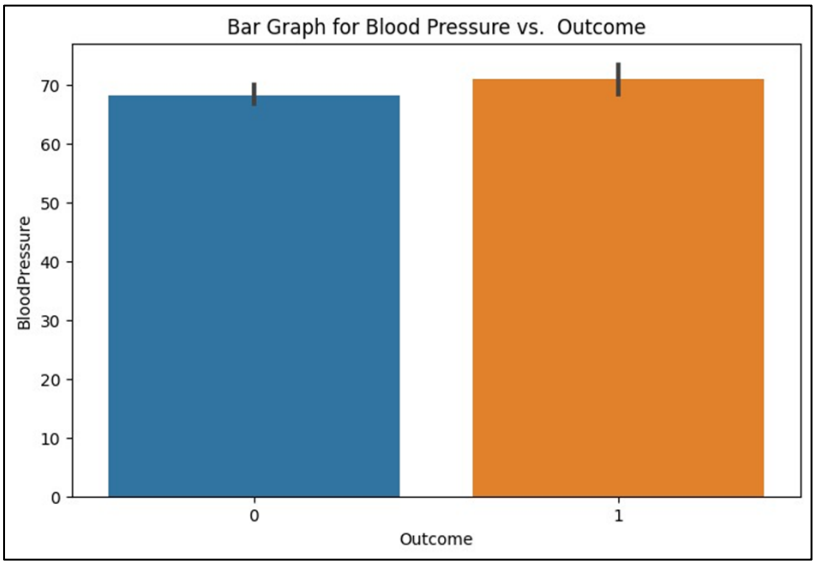

Figure 1 Representing bar graph for blood pressure

The figures below represent how different attributes play important role in diabetes mellitus. Figure 1 represents bar graph for blood pressure, Figure 2 represents histogram for age distribution, Figure 3 represents pie chart for pregnancy distribution and Figure 4 represents bar graph for insulin.

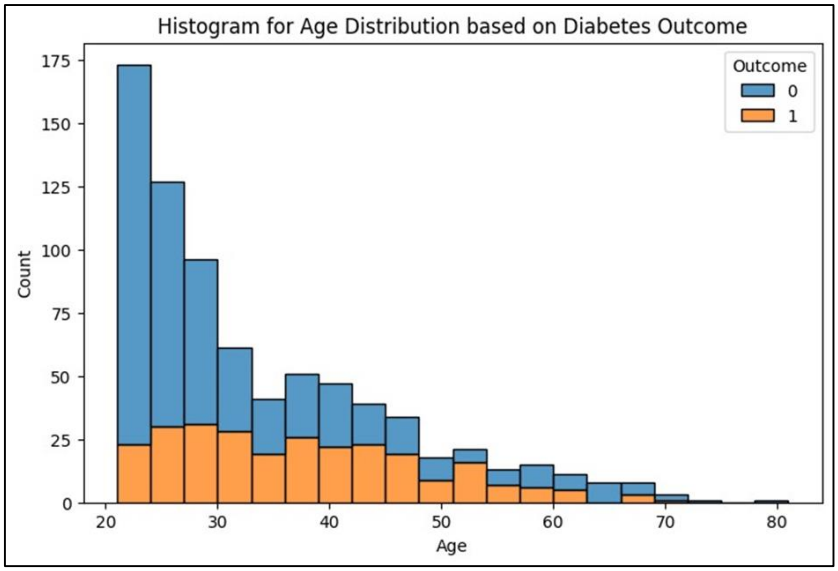

Figure 2 Representing histogram for age distribution

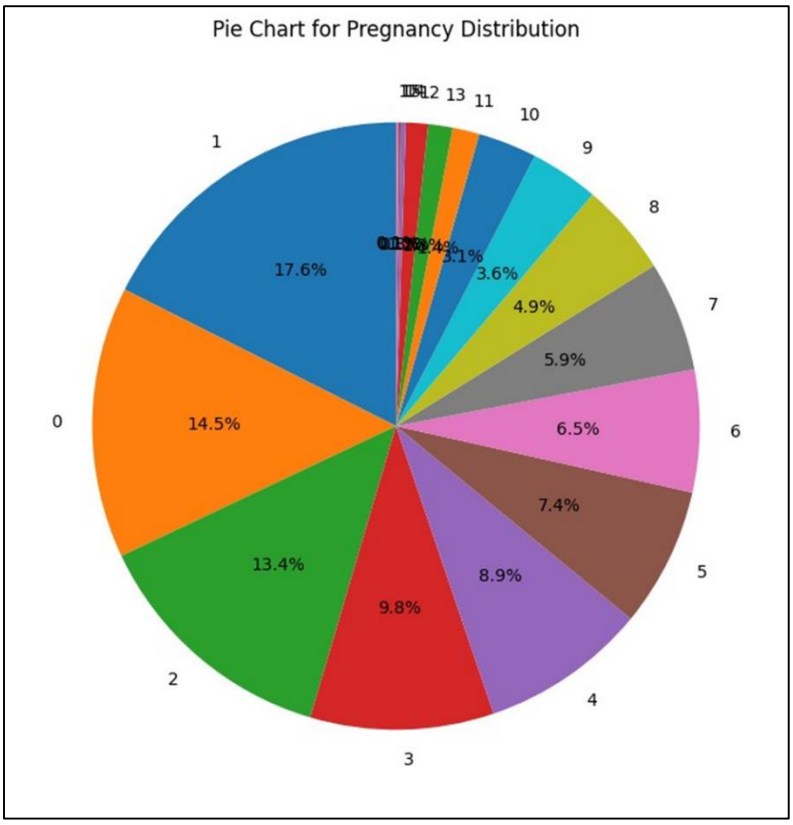

Figure 3 Representing pie chart for pregnancy distribution

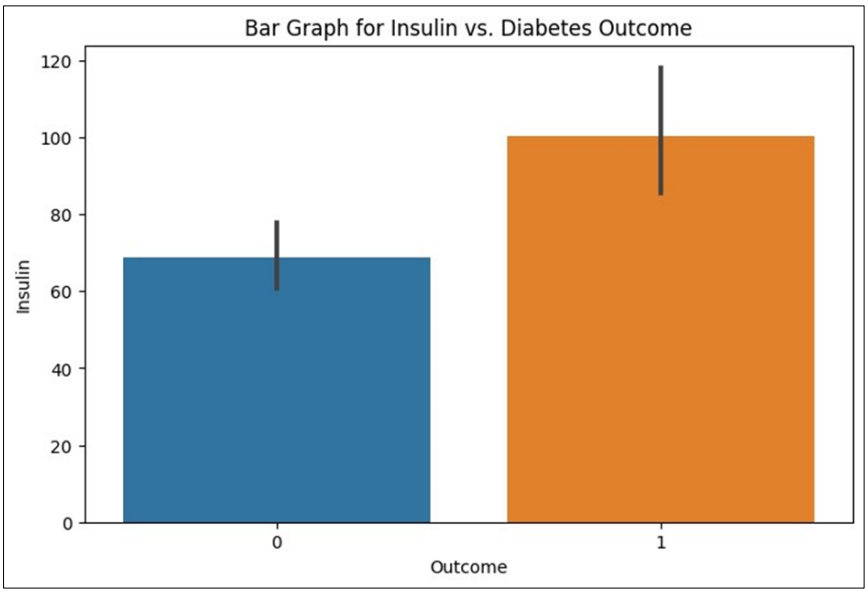

Figure 4 Representing bar graph for insulin

4. Conclusion

After a while, diabetes can be fatal if it is not properly identified or diagnosed. There have been numerous machine learning techniques discussed, beginning with various fundamental algorithms, including DTs, SVM, and LR. An evaluation of machine learning techniques for predicting diabetes early and an explanation of the use of various supervised as well as unsupervised machine learning techniques to the dataset to attain optimal accuracy used by various researchers are discussed in this paper.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Ahmed, N., Ahammed, R., Islam, M. M., Uddin, M. A., Akhter, A., Talukder, M. A., & Paul, B. K. (2021). Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering*, *2*, 229-241.

[2] Cardozo, G., Pintarelli, G. B., Andreis, G. R., Lopes, A. C. W., & Marques, J. L. B. (2022). Use of Machine Learning and Routine Laboratory Tests for Diabetes Mellitus Screening. *BioMed research international*, *2022*.

[3] Das, D., Biswas, S. K., & Bandyopadhyay, S. (2022). A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning. *Multimedia Tools and Applications*, *81*(18), 25613-25655.

[4] Firdous, S., Wagai, G. A., & Sharma, K. (2022). A survey on diabetes risk prediction using machine learning approaches. *Journal of Family Medicine and Primary Care*, *11*(11), 6929.

[5] Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & metabolic syndrome*, *13*(1), 1-22.

[6] Gouri Saria, M. K., Jee, G., Harshvardhan, G. M., Singh, V., Singh, P. K., & Workneh, T. C. (2022). Data science appositeness in diabetes mellitus diagnosis for healthcare systems of developing nations. *IET Communications*, *16*(5), 532-547.

[7] Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied computing and informatics*, *18*(1/2), 90-100.

[8] Khan, F. A., Zeb, K., Al-Rakha mi, M., Dahab, A., & Bukhari, S. A. C. (2021). Detection and prediction of diabetes using data mining: a comprehensive review. *IEEE Access*, *9*, 43711-43735.

[9] Nadeem, M. W., Goh, H. G., Ponnusamy, V., Ando Novic, I., Khan, M. A., & Hussain, M. (2021, October). A fusion-based machine learning approach for the prediction of the onset of diabetes. In *Healthcare* (Vol. 9, No. 10, p. 1393). MDPI.

[10] Saru, S., & Subashree, S. (2019). Analysis and prediction of diabetes using machine learning. *International journal of emerging technology and innovative engineering*, *5*(4).

[11] Shafi, S., & Ansari, G. A. (2021, May). Early prediction of diabetes disease & classification of algorithms using machine learning approach. In *Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021)*.