

A Study of Machine Learning Techniques for Diabetes Prediction

Neha Dhakad¹, Dr. P.K. Chande², Neha Mehra³

¹Department of Computer Engineering S.G.S.I.T.S. Indore, India, nehadhakad@gmail.com

²Ex. Prof. IIM Chairman CS Mind Indore, India, pkchandein@gmail.com

³Department of Computer Engineering S.G.S.I.T.S. Indore, India, mehra.neha40@gmail.com

Received Date : January 05, 2022

Accepted Date : January 31, 2022

Published Date : February 07, 2022

ABSTRACT

Kidney failure, heart failure, blindness, and stroke are all common complications of diabetes. When we consume, our bodies transform food into sugar or glucose. Our pancreas is meant to release insulin at that point. Insulin is a key that allows glucose to enter and be utilized for energy in our cells. The two most common types of diabetes are Type 1 and Type 2. If diabetes is detected early enough, it can be controlled. Using effective and dependable machine learning approaches to discover trends and anticipate the onset of diabetes in humans would aid in the earlier detection and treatment of the illness. Smoking, diet, stress, sleeping time, exercise, and other factors can help us to determine whether a person is prediabetic or diabetic. This study focuses on recent advances in machine learning that have a significant impact on diabetes detection and diagnosis.

Keywords: Disease, Diabetes Prediction, Machine Learning, Diabetic, Prediabetic, Health.

1. INTRODUCTION

Diabetes mellitus is one of the deadliest and most common diseases in the world. It impacts millions people all around the whole world and getting more prevalent in India. Diabetes is caused by a high-sugar level diet, as well as other bad eating habits and lifestyle choices, such as a lack of consistent physical activity. The start of the disease is also influenced by genetics. Diabetic patients will number over 592 million by 2034, according to the World Diabetes Federation. Insulin resistance is a major health issues associated with diabetes. Body sugar level does not stabilize because of a shortage of insulin in our body, hormonal imbalance and pancreatic impact cause these symptoms.

The two most common types of diabetes are Type 1 and Type 2. Diabetes type 1 is significantly more dangerous than diabetes type 2. The pancreas does not manufacture insulin in type 1 diabetes, thus the patient must rely on insulin injections to stay alive. The pancreas produces insulin in type 2 diabetes, but it is insufficient, necessitating the use of insulin-producing drugs or injections. Type 2 diabetes is the most frequent type, accounting for more than 90% of all cases.

3. Diabetes prevention and early detection are critical for global health and welfare; As a result, having an accurate diabetes diagnosis and

a mechanism for predicting the impact of a variety of events on the seriousness of diabetes becomes crucial. Use data to train the model so that it has a high degree of precision in predicting the qualities of new and unknown data or Sort them into appropriate category. The algorithms' performance is highly dependent on the dataset's characteristics. As a result, there is no universally accepted optimum algorithm for diagnosing diabetes or its effects on the patient's medications.

Diabetes datasets have been the subject of a wide range of studies. The most popular method is used to clinical data, like HbA1c levels, to forecast diabetes illness. Other factors to consider are whether or not the patient will be able to maintain a consistent glucose level, glucose level prediction, and the possibility of a critical patient being readmitted after previously been admitted. This study investigates the many sorts of diabetes predictions provided with Machine Learning Techniques.

Diabetes classification is a critical stage in diabetes prevention, detection and management in healthcare. Early detection of diabetes, on the other hand, is far more effective in diabetes management. Because a patient must see a physician regularly, the diabetes detection method appears to be laborious at first. By anticipating disease, advances in machine learning algorithms have tackled significant challenges in healthcare. In the literature, a number of diabetes prediction techniques have been proposed.

Diabetes Signs and Symptoms

The following are some of the most common diabetic symptoms:

- extreme hunger or thirst
- feeling tired lethargic
- urinating more frequently than normal
- gradual weight gain (type 2) or unexplained weight loss (type 1)
- having cuts that take a long time to heal
- blurred vision
- rashes or infections on the skin

Diabetes Diagnosis Test

A blood test will be required to determine your glucose levels. if your doctor suspects you have diabetes.

- Fasting Glucose Test
- HbA1c Test
- Oral Glucose Tolerance Test
- Random Blood Glucose Test

2. RELATED WORK

V.V.et al. [1] Has discussed, Diabetes develops when blood sugar levels increase too high. Various computerized information systems that used classifiers to predict and diagnose diabetes were documented using ANN, decision trees, Naive Bayes, and SVM algorithms. Selecting appropriate classifiers improves the system's precision and proficiency. In this decision support system, the AdaBoost algorithm is combined using Decision Stump as the basic classifier for classification. Naive Bayes, Support Vector Machine, Decision Tree can also be used as basic classifiers for the AdaBoost approach for the purpose of ensuring accuracy. The AdaBoost method is more accurate than Support Vector Machine, Naive Bayes, and Decision Tree, with an accuracy of 80.72 percent.

R.P. et al. [2] Has Diabetes Retinopathy is one of the most common causes of vision impairment in diabetic people. They investigated the performance of a variety of machine learning algorithms and confirmed their correctness for a specific data set. Diabetic Retinopathy has been predicted using machine learning methods. We employed four algorithms in this study, namely Naive Bayes, DT, KNN, and SVM, out of many available classification techniques. Each algorithm has its own set of benefits and drawbacks depending on the application. The SVM, on the other hand, had the highest accuracy among the others, according to the data.

Komi et al. [3] Describe and explain various categorization algorithms based on variables such as blood sugar, skin thickness, insulin, blood pressure, BMI, age and diabetes pedigree. Pregnancy factors were not used by the researchers to predict diabetes illness. The researchers only used a small sample of data to predict diabetes in this investigation. This paper uses five different algorithms: GMM, SVM, ANN, EM, and LR. Finally, the researchers found that ANN had a high level of accuracy in predicting diabetes.

Nilashi et al. [4] Fuzzy rules were generated using CART (Classification and Regression Tree). Prior to implementing the rule, a clustering approach (PCA and Expectation) was utilized for maximizing and noise removal. A variety of medical datasets were employed, including heart disease, diabetes and breast cancer. Develop decision-making aids for a variety of conditions, including diabetes. As a result, CART with noise reduction can be very useful and better disease prediction, has the ability to save human life from death.

Pradeep et al. [5] The performance of machine learning approaches was examined and measured in this study based on precision. As a result of their investigation, they came to the conclusion that,

accuracy of procedures differs pre-processing and post-processing. This implies that pre-processing the data set has an impact on the forecast's performance and accuracy when it comes to disease prediction. The DT method has a higher degree of precision. Using diabetes data, SVM and RF provided superior prediction after pre-processing.

Xue-Hui et al. [6] The researchers used various data mining methods to forecast diabetes using real data sets gathered by a distributed examiner in this study. This study used WEKA and SPSS for data analysis and prediction, respectively. The researchers compared 3 approaches in this study: ANN, LR, and j48. Finally, the effectiveness and precision of j48 ML techniques were determined.

Kamadi et al. [7] This study discovered issues with classification. Data minimization among the most popular challenges in classification, it plays a important role in forecast accuracy. As the researchers discovered, data should be decreased in order to achieve higher and more efficient accuracy. PCA for data pre-processing in this study comprises data minimization for improved accuracy. For prediction, Fuzzy and decision tree were utilized. Finally, it was decided that the dataset needed to be reduced to improve the results.

Shahbaz et al. [8] Using 3 machine learning techniques and the Canadian primary care sentinel monitoring Network dataset, researchers were able to forecast diabetes disease in its early stages and save lives. The results of Adaboost, Bagging, and a decision tree (J48) were compared in this study to predict diabetes. The Adaboost method performed better and was more accurate than other methods in data mining Weka tool.

Abdullah et al. [9] Health care researchers can use data mining approaches to extract new knowledge from enormous amounts of health data. Using the advancement of information technology, data mining has become a crucial asset in diabetes research, allowing for the expansion or improvement of health-care distribution, decision-making assistance, and disease monitoring.

Santhanam et al. [17] Researchers used a range of datasets in this analysis, including the Diabetes Dataset. The researchers investigated and developed models that are globally competent and capable of handling a wide range of medical datasets. However, Cross-validation assessment was not used in the classification algorithm. KNN, ANN, J48, ZeroR, Naive Bayes, filtered classifier, parameter selection, and simple cart were among the methods used in this study. In terms of accuracy, Naive Bayes surpasses these algorithms.

SUMMARY OF LITERATURE REVIEWS FOR DIABETES PREDICTION

Sn	Author	Techniques	Discoveries	Limitations
1	Komi et al. [3]	GMM, SVM, SVM, Logistic Regression and ELM	Comparison to others, ANN had the best accuracy.	In this paper the biggest flaw was its tiny sample size, which made statistical significance for any of the goals extremely challenges. The second flaw was that they did not track drug adherence directly.
2	Mehrbakhs h et al. [4]	CART, clustering Algorithm	CART developed fuzzy rules. CART, which removes noise and is based on fuzzy rules, was effective in terms of prediction.	To fully use the potential and utility of noise removal, clustering, and fuzzy rule-based algorithms for illness diagnosis, much more research needs to be done in this paper.
3	Pradeep & Dr. N. [5]	J48	It was one of the most widely used and was founded to be more accurate in this investigation. The feature selection process was used.	They did not use another classifier in this paper, such as SVM or Decision Tree, which could have improved the accuracy.
4	Xue Hui et al. [6]	Logistic Regression, KNN and J48	A comparison of the algorithm performance revealed that j48 had a high accuracy of 78.27%.	There is a lack of data filtering and diverse classifiers to improve accuracy.
5	Tao et al. [16]	KNN, Naïve Bayes, RF, DT, SVM, and logistic regression	Improvements were made to the filtering criteria. This study has a higher recall accuracy rate.	They used a small-size dataset. Using real-time dataset or big dataset, we can improve the accuracy of recall.
6	Perveen et al. [8]	j48, Bagging, and Adaboost	Adaboost was appeared to be the most accurate in this study when compared to the other options.	In an ensemble architecture different techniques like Nave Bayes, SVM, and neural networks can all be employed as foundation learners.
7	Anuja and Chitra [15]	SVM	SVM was given 78% accuracy rating.	In this paper, they did not use the feature subset selection strategy, which can improve SVM performance.
8	Santhanam and P. [17]	Genetic Algorithm, K-means, and SVM	The accuracy of a new integrated system clustering and classification algorithm has been demonstrated.	In this study, standard deviation is used to replace missing data, boxplots are used to detect outliers, SVM/PCA is used to pick features, and experiments classifiers from neural, statistical, tree and fuzzy families are used to increase accuracy.

3. DIABETES INFLUENCING PARAMETERS

Diabetes may depend on the listed below parameters:

- **Age:** The majority of diabetes diagnoses are given to adults aged 45 to 64. While there is no specific age for developing type 2 diabetes. With age, the likelihood of contracting the condition increases considerably.
- **Weight:** Many people find it difficult to lose weight and keep it off, although weight control is critical for diabetic patient health. Regular exercise can help to improve your body ability to use insulin and shed and keep weight off if you already have type 2 diabetes. Losing weight can help you manage your diabetes if you are overweight.
- **Smoking Habit:** A variety of conditions can lead to type 2 diabetes, one of which is smoking. Type 2 diabetes is 30–40 percent more likely in smokers than in nonsmokers. Furthermore, smokers with diabetes are more likely than non-smokers to struggle with insulin doses and disease management. You are more likely to get type 2 diabetes if you smoke more cigarettes. Smoking makes diabetes more difficult to manage no matter what type of diabetes you have.
- **Genetic Diabetes:** Type 2 diabetes is more closely linked to a family history of the disease. In the development of type 2 diabetes, genetics plays an important influence.
- **Exercise Habit:** Exercise aids weight loss, blood pressure regulation, muscular and bone strength, anxiety reduction, and overall well-being. Exercise decreases blood glucose levels and increases insulin sensitivity, which helps person with diabetes overcome insulin resistance.
- **Sleeping Time:** Sleep deprivation leads to a change in hormone balance, which can affect food intake and weight. This can raise your blood sugar level, making it difficult to get a good night sleep. Obesity is also linked to a lack of sleep. Obesity increases your chances of having type 2 diabetes.
- **Stress:** Blood sugar level rises as a result of stress. Stress affects sleep, which affects glucose tolerance. Blood pressure is affected by stress.
- **Diet:** Food intake timing and types should be consistent for diabetics. This assists the patient in avoiding dangerously high or low blood sugar levels. Choosing healthful foods and consuming the proper amount of food at the right time is a part of meal planning.

4. PROPOSED SYSTEM

The suggested system is based on the combination of algorithms illustrated in the block diagram below. Support Vector Machine, Decision tree, Naive Bayes, and ANN are the basic classification methods for accuracy authentication.

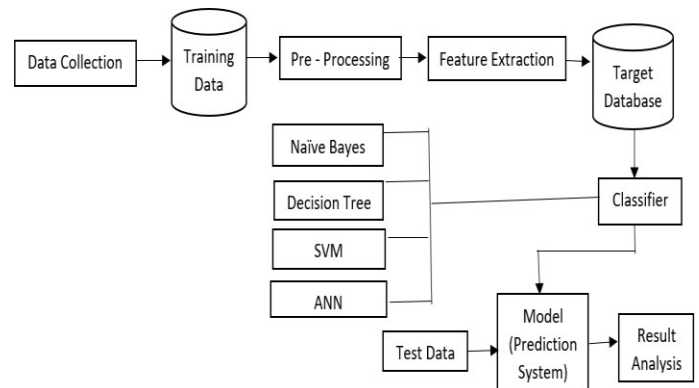


Figure 1: Diabetes Prediction System

Dataset Collection

Global dataset- Training period has come to an end. There are 768 cases and nine features in the collection (Pima Indians Diabetes Database). The following are the characteristics of the dataset:

- Glucose/sugar level
- Total number of pregnancies
- Blood Pressure Diastolic
- (BMI) Body Mass Index
- Skinfold thickness in millimetres
- Insulin level in two hours
- Pedigree function - hereditary factor
- Patient's age in years

For training and testing, a percentage split option is available. 75% of the 768 instances are for training, while 25% are for testing.

Data Preprocessing

The adjustment we do to our data before passing it on the algorithm are referred to as pre-processing. The Data To convert raw data into an useable set of information, a preprocessing strategy is utilized.

Data preprocessing is the most important phase. In healthcare data, missing values and other impurities are widespread, reducing its effectiveness. The purpose of data preprocessing is to improve the quality and efficacy of the findings obtained through the mining process. When utilizing Machine Learning Techniques on the dataset, this procedure is required for

accurate findings and appropriate prediction. Pre-processing must be done in two parts.

Missing Data Values Removal

All instances having a value of zero should be removed. There is no such thing as a value of zero. As a result, this instance isn't valid any longer. We construct a features subset by deleting non-essential features/instances, a process known as feature subset selection process which reduces data dimensionality and enables us to operate more efficiently.

Splitting Data

Data is normalized in testing and training model after it has been cleaned. When the data spit out, we use the training data set to train an algorithm while keeping test data set aside. The training model will be generated using logic and algorithms, as well as the values of the feature in the training data. The process of putting all of the traits on the same scale is known as normalization.

Training Data and Testing Data

Training data set is used to teach the model in Machine Learning, to perform a large number of actions. To train the model, detailed features are extracted from the training set. As a result, in the prototype, these structures are combined. For sentiment analysis, single words or sequences of words are collected from the tweets collection. As a result, provided training data set is appropriately labelled, model will be able to learn from the features. As a result, such information is used to test the model to verify if it is functioning properly.

Machine Learning Techniques

After the data is ready, Machine Learning Methodologies are used by us. Several classification and ensemble methods are used to predict diabetes. Methods include the following:

Decision Tree

A basic classification technique is a decision tree. DT is a monitored method of learning. When the response variable is categorical, a decision tree is utilized. The decision tree is based on a model with a tree-like structure that determines how input features are categorized. Graphs, texts, discrete, continuous, and other types of input variables are all possible. The Decision Tree Algorithm's Steps-

Step 1: Begin the tree with the root node, which contains the entire dataset, as S gets.

Step 2: Using the Attribute Selection Measure, find the best attribute in the dataset.

Step 3: Subdivide S into subgroups with the best feasible values for each attribute.

Step 4: In the decision tree, create best-attribute-containing node.

Step 5 – Utilizing the dataset's subsets established in step 3, Create a new decision tree in a recursive manner. Continue this procedure until you can no longer categorize the nodes any further, final node will be referred to as a leaf node.

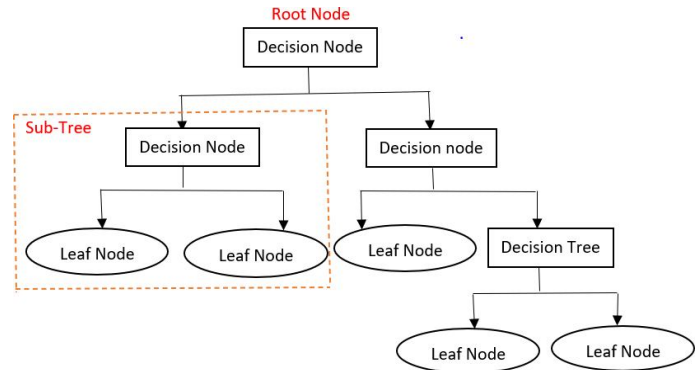


Figure 2: Decision Tree

Support Vector Machine

SVM stands for Support Vector Machine and is a supervised machine learning approach. The most often used classification method is SVM. To distinguish two classes, SVM builds a hyperplane. In high-dimensional space, it can produce a hyperplane or a succession of hyperplanes. This hyperplane can also be utilized for regression and classification. SVM may categorize and divide elements that aren't supported by data into specific classes. Separation is achieved via a hyperplane, which performs separation to the nearest training point in any class.

◆ First Category of Dataset ● Second Category of Dataset

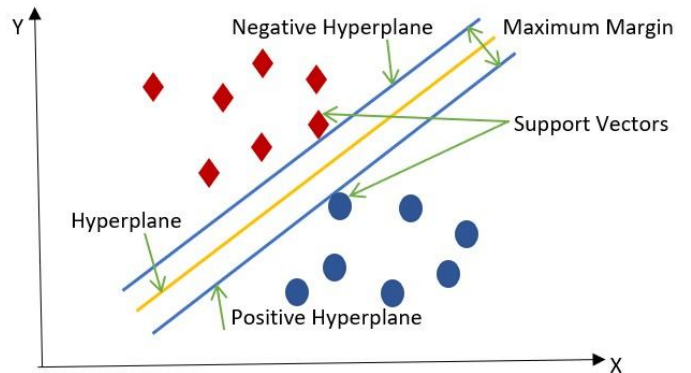


Figure 3: Support Vector Machine

The Steps of the SVM Algorithm-

Step 1: Choose the hyperplane that best divides the class.

Step 2: To determine the better hyperplane, you must calculate the Margin, which is the distance between the planes and data.

Step 3: If there is a short interval between the classes, there is a higher likelihood of miscarriage when there is a higher possibility of pregnancy, and vice versa.

Step 4: Select the class with the largest profit margin.
 Margin = Distance from positive point + Distance from negative point

Artificial Neural Network

ANN classifies incoming data into the target product via supervised learning. Because artificial neurons are constructed up of weighted interconnections that regulate the influence of input signals, supervised learning is used to identify diabetes load variables in neural networks. To begin, a neural network accumulates and recognizes a source of data to the network in the classification of diabetes. The network has received training and chooses the training procedure using a defined training dataset. After the training process, the ANN is put to the test to see how it reacts and whether it or not the disease is correctly identified.

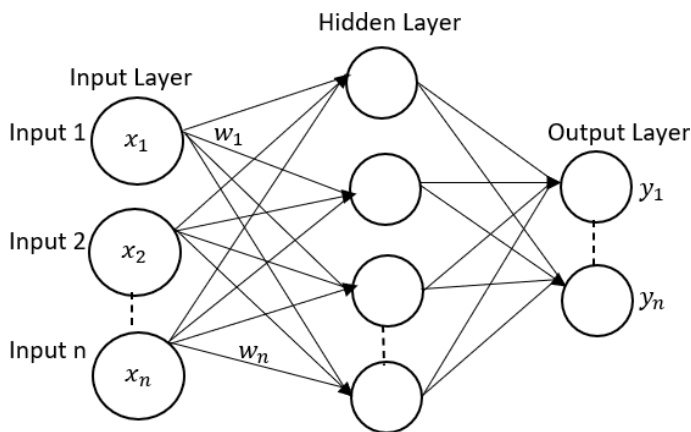


Figure 4: Artificial Neural Network

Artificial neural network, which adds a bias, computes the weighted total of the inputs. The computation is expressed using the transfer function.

$$\sum_{i=1}^n W_i * X_i + b$$

In addition to the weighted sum of the neuron's inputs, bias is an extra parameter in the Neural Network that is used to affect the output.

Model Building

This is the most crucial step, which includes the development of diabetes prediction models. For diabetes prediction, we apply the various machine learning methods outlined above.

Proposed Methodology's process

Step 1: Import the necessary libraries and the diabetic dataset.

Step 2: Remove missing data from the data by pre-processing it.

Step 3: Split the dataset in half, Training set using 80% and Testing set using 20%.

Step 4: Choose amongst Support Decision Tree, Vector Machine, K-Nearest Neighbor, and Logistic Regression as machine learning methods.

Step 5: Using the training set, create a classifier model for the stated machine learning technique.

Step 6: Test the Classifier model using the provided machine learning algorithm using the test set.

Step 7: A classifier's results are received.

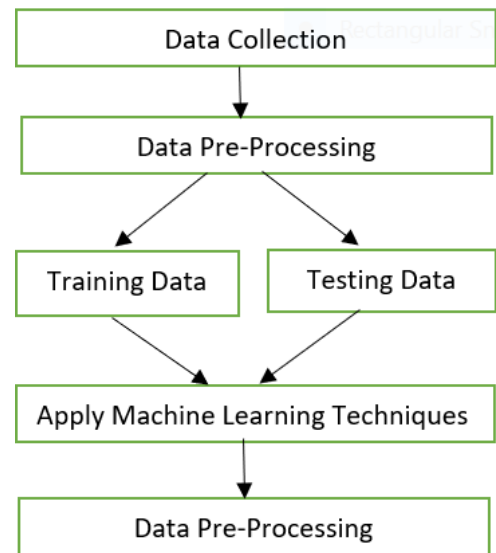


Figure 5: Overview of the Process

5. CONCLUSION

Over time, more research on diabetes diagnosis has been published. Diabetes is a condition that develop as a result of an increased level of sugar fixation in the blood. For predicting and diagnosing diabetes, many computerized information systems incorporating various classifiers were described.

Selecting valid classifiers improves the system precision and proficiency.

Finally, we have looked that how machine learning was utilized to diagnose diabetes. It is crucial to keep future exams on track, which involves the creation of diabetic detection technologies that improve performance. The anticipated future bearing could boost efficiency and broaden the range of diabetes diagnosis techniques available.

REFERENCES

1. Veena Vijayan V. And Anjali C, Prediction and Diagnosis of Diabetes Mellitus, "A Machine Learning Approach", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10- 12 December 2015 | Trivandrum.
2. Ridam Pal, Dr. Jayanta Poray, and Mainak Sen, "Application of Machine Learning Algorithms on Diabetic Retinopathy", 2017 2nd IEEE International Conference On Recent Trends In Electronics Information & Communication Technology, May 19-20, 2017, India.
3. Komi, M., Li, J., Zhai, Y., & Zhang, X. (2017, June). Application of data mining methods in diabetes prediction. In Image, Vision, and Computing (ICIVC), 2017 2nd International Conference on (pp. 1006-1010). IEEE.
4. Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An Analytical Method for Diseases Prediction Using Machine Learning Techniques. *Computers & Chemical Engineering*.
5. Pradeep, K. R., & Naveen, N. C. (2016, December). Predictive analysis of diabetes using J48 algorithm of classification techniques. In Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on (pp. 347-352). IEEE.
6. Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung journal of medical sciences*, 29(2), 93-99.
7. Kamadi, V. V., Allam, A. R., & Thummala, S. M. (2016). A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach. *Applied Soft Computing*, 49, 137-145.
8. Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, 82, 115-121.
9. Aljumah, A. A., Ahamad, M. G., & Siddiqui, M. K. (2013). Application of data mining: Diabetes health care in young and old patients. *Journal of King Saud University-Computer and Information Sciences*, 25(2), 127-136.
10. Saravananathan, K., & Velmurugan, T. (2016). Analyzing Diabetic Data using Classification Algorithms in Data Mining. *Indian Journal of Science and Technology*, 9(43).
11. Prajwala, T. R. (2015). A comparative study on decision tree and random forest using the R tool. *International journal of advanced research in computer and communication engineering*, 4, 196-1.
12. Hashi, E. K., Zaman, M. S. U., & Hasan, M. R. (2017, February). An expert clinical decision support system to predict disease using classification techniques. In *Electrical, Computer and Communication Engineering (ECCE), International Conference on* (pp. 396-400). IEEE.
13. AlJarullah, A. A. (2011, April). Decision tree discovery for the diagnosis of type II diabetes. In *Innovations in Information Technology (IIT), 2011 International Conference on* (pp. 303-307). IEEE.
14. Guo, Y., Bai, G., & Hu, Y. (2012, December). Using Bayes network for prediction of type-2 diabetes. In *Internet Technology And Secured Transactions, 2012 International Conference for* (pp. 471-472). IEEE.
15. Kumari, V. A., & Chitra, R. (2013). Classification of diabetes disease using support vector machine. *International Journal of Engineering Research and Applications*, 3(2), 1797-1801.
16. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. *International journal of medical informatics*, 97, 120-127.
17. Santhanam, T., & Padmavathi, M. S. (2015). Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. *Procedia Computer Science*, 47, 76-83.
18. Kandhasamy, J. P., & Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51.
19. kumar Dewangan, A., & Agrawal, P. (2015). Classification of Diabetes Mellitus Using Machine Learning Techniques. *International Journal of Engineering and Applied Sciences*, 2(5), 145-148.
20. Nai-arun, N., & Moungrmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. *Procedia Computer Science*, 69, 132-142.
21. S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using a soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, 2021.
22. A.K., Dewangan, and P., Agrawal, "Classification of Diabetes Mellitus Using Machine Learning Techniques," *International Journal of Engineering and Applied Sciences*, vol. 2, 2015.
23. Md. Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus". *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 7-9 February 2019.
24. S. Gupta, H. K. Verma, and D. Bhardwaj, "Classification of diabetes using Naïve Bayes and support vector machine

- as a technique,” *Operations Management and Systems Engineering*, Springer, Singapore, pp. 365–376, 2021.
25. M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, “Classification and prediction of diabetes disease using machine learning paradigm,” *Health Information Science and Systems*, vol. 8, no. 1, pp. 7–14, 2020.
 26. M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, “Likelihood prediction of diabetes at an early stage using data mining techniques,” in *Computer Vision and Machine Intelligence in Medical Image Analysis*, pp. 113–125, Springer, Singapore, 2020.
 27. D. Yang, Y. Yang, Y. Li, and R. Han, “Physical exercise as therapy for type 2 diabetes mellitus: from mechanism to orientation,” *Annals of Nutrition and Metabolism*, vol. 74, no. 4, pp. 313–321, 2019.
 28. K. Saravananathan and T. Velmurugan, “Analyzing diabetic data using classification algorithms in data mining,” *Indian Journal of Science and Technology*, vol. 9, no. 43, pp. 1–6, 2016.
 29. D. Sisodia and D. S. Sisodia, “Prediction of diabetes using classification algorithms,” *Procedia computer science*, vol. 132, pp. 1578–1585, 2018.
 30. Kimi Vyas, P. K. Chande, P. Bamne, “DiaM-Integrated Mobile-Based Diabetes Management”, vol. 9. No.4 4, April 2021.
<http://www.warse.org/IJETER/static/pdf/file/ijeter24942021.pdf>