# Evaluation on Diabetes Care Using Machine Learning

Lalit Mohan
*CSE Department*
*Graphic Era Hill University*
Bhimtaal, India
lalitmohan@gehu.ac.in

Priyanka Suyal
*CSE Department*
*COER University*
Roorkee, India
priyankasuyal48@gmail.com

Ravindra Singh Koranga
*CSE Department*
*Graphic Era Hill University*
Bhimtaal, India
333koranga@gmail.com

Bhawna Tewari
*CSE Department*
*Graphic Era Hill University*
Bhimtaal, India
bhawnatewari.200121157@gehu.ac.in

Amit Mittal
*Department of Allied Science*
*Graphic Era Hill University*
Bhimtaal, India
amitforestry26@gmail.com

2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10306462

***Abstract*— Diabetes is a chronic condition defined by elevated blood sugar levels. Machine learning (ML) has become more and more applied in the prediction and management of diabetes. A significant public health issue, diabetes affects 425 million persons globally. Diabetes consequences like cardiovascular disease, renal failure, and blindness can be avoided or delayed with early diagnosis and effective care of the condition. Additionally, ML has been applied to increase the precision of diabetes diagnosis testing. For instance, when a neural network was used to analyze the results of an oral glucose tolerance test, the model outperformed conventional diagnostic techniques in terms of sensitivity and specificity. Additionally, ML has been used to personalize diabetic medication, track the progression of disease, and forecast the likelihood of complications.**

***Keywords*— *Machine Learning, Supervised, Unsupervised, reinforcement.***

## I. INTRODUCTION

Diabetes is a serious medical condition that develops when the level of glucose, in the human blood becomes too high. Glucose is the primary source of energy that comes from the foods we eat. If left untreated, high blood sugar levels can lead to a range of complications such as kidney disease and blindness. Diabetes are classified as Type 1 Diabetes, Type 2 Diabetesand Gestational.

Type 1 diabetes

Type 1 diabetes is a condition in which the body does not produce enough insulin, and the immune system attacks and destroys the cells in the pancreas responsible for insulin production. This form of diabetes is commonly diagnosed in children and young adults, but can occur at any age. Individuals with type 1 diabetes require daily insulin injections to manage their blood sugar levels.

Type 2

The commonly observed variant is Type 2 diabetes and often affects middle-aged and older individuals. This type of diabetes occurs when the body either does not properly use insulin or does not produce enough insulin to meet its needs.

It is a type of diabetes that can develop during ppregnancy in women who don't already have diabetes. Every year, 2% to 10% of pregnancies in the United States are affected by gestational diabetes.

During pregnancy, the placenta can inhibit the absorption of insulin in the body's cells, resulting in high blood sugar levels. This condition is known as gestational diabetes and typically resolves after pregnancy. However, babies of such women have high chance of developing type 2 diabetes.

The healthcare industry generates a wealth of valuable data on the various risks associated with different types of diabetes, including electronic medical records., and diabetes information this can help in predictive analysis and determination for reducing risk management. Early detection and diagnosis of diabetes are possible due to recent advancements in the technological development of Internet of Things , AI, and blockchain in the current medical maintenance. the revolution of intelligence analysis methods There are several data mining as well as machine learning methods to manage and extract insights from large volumes of data, such as electronic medical records. These techniques have significant potential for predictive analysis and knowledge extraction [12]. Machine learning and deep learning are both AI-based techniques that offer unique strengths and applications within the field..Deep learning is used in analysing patterns and extracting rules whereas ml has the potential of improving efficiency and reducing the expenses associated with healthcare treatment in the system. ml is also significant for learning and automating machine and pattern recognition[11].

In the medical field, classification strategies are widely utilized to categorize information into various classes. according to some constraints compared to an individual classifier.

Support vector machine (SVM):

SVM, as a common group of supervised machine learning models, is utilized in classification tasks involving two classes. Support vector machines (SVMs) are an example of neural network.

A study has compared the implementation of several machine learning techniques on a diabetes dataset. The results showed that Random Forest has the highest classification accuracy of 99.84% and the Support Vector Classifier has the highest precision of 99.64%. However, other models such as KNN, ANN, SVM with the RBF kernel, and Multifactor Dimensionality Reduction performed moderately on the Kernel Entropy Component Analysis trained dataset [13].

The evaluation of each model's performance was cconducted by considering several parameters, including recall, accuracy, F1 score, precision, and ROC-AUC. Among the algorithms tested, SVM-linear achieved the maximum accuracy of 89% and maximum precision of 88%, whereas K-Nearest Neighbor exhibited the highest recall of 90% and F1 score of 88%. Based on these results, K-Nearest Neighbor and SVM-linear emerged as the preferred algorithms, making them the optimal choices to diagnose diabetes. [14].

## II. Machine Learning Techniques for Diabetes Detection

### A. *Supervised Learning*

Algorithms for supervised learning utilize direct feedback to make predictions. Common supervised learning algorithms include SVM, DT, KNN, LR, ANN, and NB. The primary goal of allocation systems is to accurately forecast the risk of diabetes sufferers.

In a study using the PIDD database ML Algorithms like SVM, Decision Tree, and Naive Bayes were applied. The Naive Bayes algorithm achieved the highest accuracy of 76.30%. In another study using the Diabetes Data Set, AdaBoost Classifier outperformed other techniques with the highest accuracy of 98.8%.

A diabetes detection method using machine learning techniques was proposed in another study. Support Vector Machine and Random Forest algorithms were used, with the Random Forest algorithm achieving better accuracy compared to SVM.

Various ML algorithms such as Decision Tree, ANN, Naive Bayes, and SVM were used to predict the risk of diabetes in patients in multiple studies. SVM and ANN algorithms had the highest accuracy in one study, while Random Forest had the highest accuracy in another.

In another study, SVM with a linear kernel was found to have the highest accuracy for classifying diabetes. Another study proposed the use of the Random Forest algorithm to perform early prediction of diabetes in patients with higher accuracy. The performance of SVM was analysed for different kernels, and the best kernel was selected and used for prediction.

In summary, supervised learning algorithms such as SVM, DT, KNN, LR, ANN, and NB have been used to predict the risk of diabetes in patients. Different studies have shown varying results regarding the accuracy of these algorithms, with some studies finding Random Forest or Naive Bayes to have the highest accuracy. The choice of algorithm and kernel may depend on the specific dataset and research question. .

TABLE I. TABULATION OF SUPERVISED LEARNING ALGORITHMS

| S. N | Ref. No. | Findings | Best Algorithm |
|---|---|---|---|
| 1 | [4] | Support Vector Machine and Random Forest, two machine learning methods used for diabetes prediction, are combined in the suggested model.. | Random Forest |
| 2. | [3] | Different supervised machine learning methods were evaluated to determine the most suitable algorithm for diabetes prediction. | AdaBoost classifier |
| 4. | [5] | The critical factors for the development of diabetes were identified and multiple machine learning algorithms, including logistic regression (LR), support vector machine (SVM), and random forest (RF), were compared to predict diabetes.. | Random Forest |
| 5. | [6] | The study employed two classification techniques from machine learning and compared their results using various statistical tests. | Random Forest |
| 6. | [7] | The study applied various kernel functions to SVM. | SVM with linear kernel |
| | | The performance | SVM |

| | | | |
|---|---|---|---|
| 7. | [8] | of Random Forest algorithms was studied and evaluated on various measures. The aim was to develop a system that can accurately predict diabetes at an early stage for patients, utilizing machine learning techniques. | with linear kernel |
| 8. | [9] | The study utilized various families of algorithms, including logistic regression, support vector machines with linear and nonlinear kernel. | Random Forest and SMO |
| 9. | [10] | The author evaluated Mutliple datamining classification techniques using MatLab. | ANN with 89% |

*B. Unsupervised Learning*

Algorithms for unsupervised learning, such as cluster analysis and Principal Component Analysis (PCA), can identify hidden patterns in data without using any input. PCA can eliminate strongly correlated features by using the covariance matrix, Self-Organized Model (SOM), Apriori algorithm, Singular value decomposition and PCA are popular unsupervised algorithms. To detect and pre-process outliers clustering algorithm can be used. In a work published in [13], the authors developed a reinforcement learning model that recommends the number of oral antidiabetic medications and insulins for people with type 2 diabetes. The model was trained on data from a patient registry and was shown to improve haemoglobin achievement by 15%. In a study , the Boruta algorithm was used to identify substantially linked features of BMI and Plasma glucose, and Gradient Boosting was found to have a prediction accuracy of 86% in regression and classification tasks. However, the study in [ found that SVM with K-Mean on the complete dataset had poor performance compared to other algorithms, but combining SVM and decision trees for data resolution produced ssuperior results. The following papers have been thoroughly examined in relation to unsupervised learning and were included in the study. in Table II.

TABLE II.

TABLE III. TABULATION OF UNSUPERVISED LEARNING ALGORITHMS

| S. N | Ref. No. | Findings | Best Algorithm |
|---|---|---|---|
| 1. | [13] | The accuracy of the predictive model developed using the K-means algorithm was marked down at 78%. The researcher also intends to implement other clustering algorithms in R tool and compare the accuracy and sensitivity of all the algorithms. | K means clustering algorithm |
| 2. | [15] | The linear kernel achieved the highest accuracy, with up to 83%, while integrated technology achieved an accuracy of 82%. Furthermore, the combination of SVM and decision tree resulted in even more accurate predictions. | SVM & Decision tree |

*C. Reinforcement Learning*

Reinforcement learning (RL), a potential machine learning training technique that rewards good behaviours and penalises unfavourable ones, has gained popularity in recent years. Real-world (RL) models are able to act, take in their environment, and learn from their failures. In this study, the authors used RL with a deep Q network to learn from the dataset. The amount of oral antidiuretic medications and insulin is advised by the RL model to enhance long-term glucose control. The trials demonstrated that the RL model could boost haemoglobin attainment rates, emonstrating the algorithm's capacity to learn effective prescription patterns .

Traditional RL algorithms need properly selected feature representations, but they nevertheless offer an intelligent, individualised, and superior method for estimating insulin delivery. Recent patents on optimising patient treatment recommendations using RL-based insulin dose estimation and RL in combination with a neural network have been released. Few articles have used this family of algorithms in the BG control problem, indicating that RL is still a relatively new strategy in the field of diabetes. During their comprehensive assessment of the research, the authors of this study discovered traits that exactly fit the trends in the literature.

The following papers have been thoroughly examined in the context of reinforcement learning and were included in the representation Table III.

TABLE IV. TABULATION OF REINFORCEMENT LEARNING ALGORITHMS

| S. N | Ref. No. | Findings | Best Algorithm |
|---|---|---|---|
| 1. | [1] | We have observed that treatments adhering to our RL algorithms's recommendations has shown improvement in HbA1c levels along with improved result in glucose management after 12 months. Additionally, our experimental findings indicate that our RL approach can effectively learn the optimal medication, leading to successful glycemic control. | deep Q network |
| 2. | [2] | RL algorithms have emerged as an intelligent, personalized, and optimal solution for determining insulin delivery. In this context, it is noteworthy to introduce a current patent that focuses on estimating insulin dosage using RL, and another that combines Reinforcement Learning and deep learning to provide better treatment exhortations for patients, with diabetes being a practical application example. | IHD |

## III. CONCLUSION

Machine learning (ML) algorithms applied in tracking and forecasting diabetes has demonstrated promise in recent years. To recognise people having diabetes, machine learning (ML) can analyse information about patients, enabling an earlier diagnosis and improved disease treatment. This study reviews and assesses a variety of cutting-edge methods for predicting and detecting diabetes. In order to identify diabetes early, supervised learning methods, including N-Bayes, SVM, and decision trees, have been used. In this area, decision trees have demonstrated promise, and combining them with unsupervised methods like PCA and K-Mean can increase prediction precision and accuracy. Additionally showing promise in effectively recognising and analysing diabetes are K-Mean and SVM. Furthermore, deep learning-based algorithms like ANN and CNN can improve diabetic patient outcomes.

## REFERENCES


[1] Zhuo Liu, Linong Ji, Xuehan Jiang, Wei Zhao," A Deep Reinforcement Learning Approach for Type 2 Diabetes Mellitus Treatment", 2020 IEEE International Conference on Healthcare Informatics (ICHI) | 978-1-7281-5382-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICHI48887.2020.9374313

[2] Miguel Tejedor, Ashenafi Zebene Woldaregay, Fred Godtliebsen," Reinforcement Learning Application in Diabetes Blood Glucose Control: A Systematic Review" Tromsø Research Foundation, University of Tromsø-The Arctic University of Norway

[3] Aishwarya Mujumdar, V Vaidehi,Diabetes Prediction using Machine Learning Algorithms,Procedia Computer Science,Volume 165,2019,Pages 292-299,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.01.047.

[4] A. S. Alanazi and M. A. Mezher, "Using Machine Learning Algorithms For Prediction Of Diabetes Mellitus," 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, 2020, pp. 1-3, doi: 10.1109/ICCIT-144147971.2020.9213708.

[5] D. Dutta, D. Paul and P. Ghosh, "Analysing Feature Importances for Diabetes Prediction using Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2018, pp. 924-928, doi: 10.1109/IEMCON.2018.8614871.

[6] V. Mounika, D. S. Neeli, G. S. Sree, P. Mourya and M. A. Babu, "Prediction of Type-2 Diabetes using Machine Learning Algorithms," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 127-131, doi: 10.1109/ICAIS50930.2021.9395985.

[7] G. A. Pethunachiyar, "Classification Of Diabetes Patients Using Kernel Based Support Vector Machines," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-4, doi: 10.1109/ICCCI48352.2020.9104185.

[8] P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 367-371, doi: 10.1109/ICCMC.2019.8819841.

[9] B. Anishfathima, P. Gautham, B. Gowri Mahalakshmi and S. Jahangir Jamadar, "Smart Architecture for Diabetic Patients Using Machine Learning," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1544-1548, doi: 10.1109/ICACCS51430.2021.9441985.

[10] M. Komi, Jun Li, Yongxin Zhai and Xianguo Zhang, "Application of data mining methods in diabetes prediction," 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, 2017, pp. 1006-1010, doi: 10.1109/ICIVC.2017.7984706.

[11] D. Shetty, K. Rit, S. Shaikh and N. Patil, "Diabetes disease prediction using data mining," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8276012.

[12] Y A. Stokes, S.H. Preston, Deaths attributable to diabetes in the Unite States: comparison of data sources and estimation approaches, PLoS One 12 (1) (2017) e0170219.

[13] Sujatha, Christy. (2018). Building Predictive Model For Diabetics Data Using K Means Algorithm..

[14] Cai, Yun Lei et al. "A KNN Research Paper Classification Method Based on Shared Nearest Neighbor." NTCIR Conference on Evaluation of Information Access Technologies (2010).

[15] N. I. Alghurair, "A Survey Study Support Vector Machines and K- MEAN Algorithms for Diabetes Dataset" . Academic Journal of Research and Scientific Publishing, vol.2,pp. 14-25, 2020.

[16] L. Mohan, J. Pant, P. Suyal and A. Kumar, "Support Vector Machine Accuracy Improvement with Classification," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 2020, pp. 477-481, doi: 10.1109/CICN49253.2020.9242572.