rice 2023
Proceedings of the Eighth International Conference on Research in Intelligent Computing in Engineering pp. 1–6
DOI: 10.15439/2023R43
ISSN 2300-5963 ACSIS, Vol. 38

Applications of Machine Learning for Diabetes Prediction: A Comprehensive Review

Nayeem Ahmed
Department of Computer Science & Information Technology
Maulana Azad National Urdu University
Hyderabad, India
naeemahmed0410@gmail.com.

Syed Imtiyaz Hassan
Department of Computer Science & Information Technology
Maulana Azad National Urdu University
Gachibowli Hyderabad, India
s.imtiyaz@gmail.com.

Md. Zair Hussain
Polytechnic Hyderabad
Maulana Azad National Urdu University
Gachibowli Hyderabad, India
zair@manuu.edu.in

***Abstract*—The use of machine learning techniques has drawn more attention due to its potential to improve early identification and intervention in diabetes, a critical global health concern. This article offers an extensive overview of the various machine learning algorithms used in diabetes prediction, including ensemble techniques, logistic regression, support vector machines, decision trees, and neural networks. The research closely examines how these algorithms make use of a variety of data sources, including wearable sensor data, electronic health records, clinical data, and genetic information. The report also emphasizes the difficulties that these applications face, including as interpretability, model integration into clinical procedures, and ethical issues. This review elucidates the significant influence of machine learning on diabetes prediction, paving the way for more useful risk assessment, individualized therapies, and improved patient outcomes. It does this by thoroughly examining recent studies and their conclusions.**

***Index Terms*—Machine Learning, Data Mining, Diabetes Prediction, Support Vector Machine, Neural Networks**

I. Introduction

Diabetes has emerged as a global health challenge, with its prevalence reaching alarming levels worldwide. As illustrated in figure.1, the International Diabetes Federation estimates that 463 million individuals worldwide had diabetes in 2019 and that 700 million will have the disease by 2045 [1]. Diabetes causes serious consequences including cardiovascular disease, renal failure, and blindness, placing a significant strain on patients, healthcare systems, and society at large [2].

Early prediction and intervention play a crucial role in managing diabetes effectively and reducing its impact on individuals' health outcomes [3]. Traditional approaches to diabetes prediction have relied on clinical risk scores and biomarkers, but they often lack accuracy and fail to capture the complexity of the disease. A paradigm change in diabetes prediction has been brought about by the development of machine learning (ML) techniques, which have the potential to enhance the precision, effectiveness, and personalized character of predictive models [4].

A substantial corpus of research has been done in the last ten years on using machine learning algorithms to predict diabetes. This research has utilised several ML methods, such as Logistic Regression [5], Decision Trees [6], Support Vector Machines [7], Neural Networks [8], and deep learning models, to create prediction models that can precisely identify people at risk of getting diabetes. By utilizing diverse data sources such as electronic health records (EHRs), genomic data, wearable devices, and behavioral data, ML models can capture the complex interplay of factors contributing to diabetes onset.

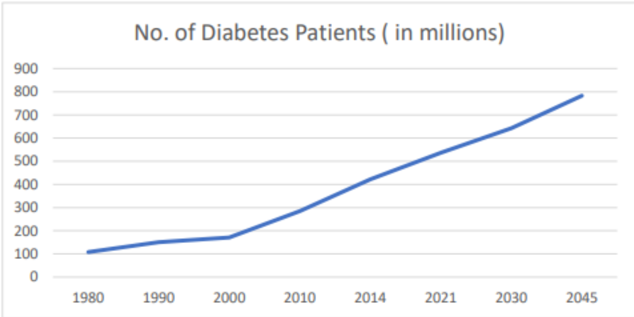

Figure.1: No of Diabetes cases worldwide [1]

The skills of machine learning in predicting diabetes have been demonstrated in a number of significant research studies. As an example, researchers in [4] successfully created a prediction model to evaluate the risk of diabetes by utilizing different classifiers on the PIMA dataset. The effectiveness and accuracy of data mining and machine learning techniques in reducing risk variables were demonstrated in this study. A noteworthy study in [9] combined genomic and clinical data, using deep learning techniques to improve prediction accuracy and reveal the underlying genetic architecture of diabetes.

These key studies and others have paved the way for advancements in diabetes prediction, showcasing the potential of machine learning techniques to revolutionize clinical decision_making and improve patient outcomes. ML models offer the ability to integrate large-scale, heterogeneous data, uncover hidden patterns, and generate accurate risk predictions tailored to individual patients.

Our goal in this large review paper is to provide a thorough summary of the most recent machine learning-based research on diabetes prediction. We will examine the methodology used in these investigations, assess their contributions to the literature, and highlight the most important outcomes and difficulties found. To advance the science of diabetes prediction and direct the creation of more precise and clinically useful ML models, we will synthesize the ex-

©PTI 2023

isting literature to find gaps and opportunities for future study.

The subsequent sections of the paper are structured as follows: Section 2 introduces the machine learning algorithms utilized for diabetes prediction. Section 3 discusses the data sources and features utilized in diabetes prediction. Section 4 focuses on the applications of machine learning in diabetes prediction. The impacts, challenges, and future directions in this field are addressed in Section 5. Finally, Section 6 concludes the paper with a conclusion.

II. Machine Learning Algorithms for Diabetes Prediction

Machine learning algorithms have been extensively applied to diabetes prediction, offering valuable insights and improved accuracy in identifying individuals at risk of developing the disease. Within this section, we will delve into the intricacies of diverse machine learning algorithms, exploring their specific applications in diabetes prediction.

A. Logistic Regression

Diabetes can be predicted well using the well-liked method for binary classification problems, logistic regression [5]. This strategy effectively replicates the relationship between independent variables and the probability of a particular outcome. In studies predicting diabetes, researchers have used logistic regression models, frequently integrating clinical and genetic variables.

For instance, researchers [10] used logistic regression models in a study to foretell the onset of diabetes. They gathered a wide range of clinical characteristics, including age, BMI, blood pressure, and genetic markers. They discovered that the addition of genetic markers considerably improved the predictive performance of the model by studying the data from a large cohort of patients. The study demonstrated how logistic regression may use a mix of clinical and genetic data to detect diabetes risk early. This demonstrates the algorithm's potential for early diabetes onset prediction.

B. Decision Trees and Random Forests

Decision tree algorithms, such as C4.5 and CART, have been widely used in diabetes prediction due to their interpretability and ability to handle both numerical and categorical data [6]. Decision trees recursively split the data based on features to create a tree-like structure, allowing for easy interpretation of the prediction process. Random forests, an ensemble method based on decision trees, combine multiple decision trees to improve the model's performance and robustness.

Researchers used decision tree algorithms to identify risk variables related with type2 diabetes by examining a large dataset of electronic health records in their work, which was published in [11]. They built decision tree models to identify critical factors linked with diabetes onset by analyzing several clinical variables such as BMI, fasting glucose levels, and family history. Their findings provide light on the fundamental elements that contribute to the development of diabetes.

Random forests have also been utilized to increase the model's accuracy and generalization capability in diabetes prediction. The authors of [12] used a mix of clinical and genetic factors to predict diabetes using random forest models. Their study found that adding both types of data enhanced prediction ability when compared to models that just used clinical variables. The random forest method shows promise in capturing the complicated interactions between multiple risk factors and the onset of diabetes.

C. Support Vector Machines (SVM)

Support Vector Machines (SVM) are robust supervised learning algorithms employed for classification tasks, including diabetes prediction. The primary objective of SVM is to determine an optimal hyper plane that effectively separates different classes by maximizing the margin between them [7].

Researchers in [13] employed SVM for diabetes prediction, utilizing a combination of clinical measurements, genetic markers, and lifestyle factors. Their study demonstrated the effectiveness of SVM in accurately classifying individuals at risk of developing diabetes. By integrating diverse data sources, including genetic and lifestyle factors, their SVM model achieved high prediction accuracy.

D. Neural Networks

Neural networks, especially deep learning models as referenced in [8], have garnered substantial attention in diabetes prediction owing to their capability to discern intricate patterns from high-dimensional data. Deep learning models consist of multiple layers of interconnected artificial neurons that can extract and process features automatically.

The authors of [14] suggested DiaNet, a revolutionary deep learning network for predicting diabetes from retinal pictures. DiaNet has an outstanding accuracy of more than 84% in detecting important retinal regions and distinguishing the Qatari diabetic cohort from the control group. The study found that retinal images include predictive markers for diabetes and other co morbidities, implying that retinal images could be used in clinical diagnosis in the future.

E. Ensemble Methods

Ensemble methods combine multiple weak classifiers to create a strong predictive model. AdaBoost [15] and Gradient Boosting [16] are popular ensemble techniques used in diabetes prediction. These methods iteratively train weak classifiers and assign higher weights to misclassified instances, focusing on the difficult samples.

The researchers introduce eDiaPredict in [17], an ensemble-based system for diabetes prediction that uses a variety of machine learning methods, including Support Vector Machine, Neural Network, Random Forest, XGBoost, and Decision tree. This technique remarkably achieves a 95% accuracy rate when applied to the PIMA Indian diabetes dataset. This emphasizes the value of effective machine learning algorithms in predicting and identifying severe situations in diabetes patients early on.

The items in table 2.1 provide an example of the wide range of machine learning techniques used in diabetes prediction. The ability to correctly identify people who are at risk of getting diabetes has significantly improved thanks to the use of ensemble methods, logistic regression, decision trees, support vector machines, neural networks, and ensem-

TABLE II. EXISTING WORKS ON DIABETES PREDICTION USING MACHINE LEARNING ALGORITHMS

Study	Algorithm	Dataset	Features	Accuracy (%)	Key findings
[10]	Logistic Regression	HER data	Clinical Genetic	82.5	Inclusion of genetic makers significantly improves prediction performance.
[11]	Decision Tree	Electronic health records	BMI, Fasting glucose levels, Family history	75.3	Identified key risk factors associated with type 2 diabetes onsets.
[12]	Random Forest	Clinical and genetic data	Clinical Genetic	87.9	Incorporating genetic features enhances prediction accuracy compared to using only clinical data.
[13]	Support Vector Machine	Clinical Measurements, Genetic markers, Lifestyle factors	Clinical Genetic Lifestyle	81.2	SVM accurately classifies individual's diabetes by integrating diverse data sources.
[14]	Neural Networks	Retinal Images	Retinal Images	91.6	Deep learning model using CNN architecture achieves high accuracy in non-invasive diabetes detection.
[17]	Ensemble Methods	Clinical & demographic data	Clinical Demographic	95	Logistic Regression is an efficient algorithm for prediction models with data preprocessing normalization and ensemble techniques enhancing overall performance.
[18]	Logistic Regression	Prima Indians Diabetes Database (PIDD)	Clinical Genetic	95.20	Logistic regression model provides accurate prediction using combined clinical and genetic features.
[19]	Random Forest	Prima Indians Diabetes Database	Clinical Genetic	83.67	When predicting diabetes, RF performed better than the deep learning and SVM techniques.
[20]	Support Vector Machine	Prima Indians Diabetes Database	Clinical Lifestyle	98.7	The proposed architecture using K-means clustering and SVM achieved an accuracy of 98.7% in predicting diabetes patients.
[21]	Artificial Neural Networks	Kurdistan region dataset	Genetic	91	The error rate decreased during training, indicating improved prediction accuracy based on network design.
[22]	Ensemble Methods (AdaBoost)	Diabetes UCI dataset	Clinical	98	Diabetes prediction using the AdaBoost M1 ensemble algorithm has a 98% accuracy rate.

ble approaches. These algorithms, when used in conjunction with sensible feature selection and model optimization, have the potential to enhance clinical decision-making and enable tailored treatments for the management of diabetes.

III. Applications of Machine Learning in Diabetes Prediction

Machine learning has been applied to various specific applications in diabetes prediction, offering valuable insights and potential clinical implications. This section, discusses existing studies that have utilized machine learning for applications such as early detection, risk stratification, and personalized treatment in the context of diabetes prediction.

The ability to quickly intervene and prevent complications depends on the early identification of diabetes. Machine learning algorithms have shown promise in identifying persons at risk of diabetes before clinical symptoms appear [23].

Based on a person's likelihood of getting diabetes or issues associated to diabetes, machine learning models can help group people into various risk categories. This enables tailored treatment regimens and targeted actions [24].

Building models that can analyze patient-specific data and generate specialized predictions for diabetes diagnosis, management, and therapy is a key component of personalized treatment for diabetes prediction using machine learning. Predictive models are created using machine learning algorithms that are trained on a variety of patient variables, including medical history, genetic information, lifestyle factors, and biomarkers [25].

A. Integration of Machine Learning Model into the Diabetes Clinical Workflow

Machine learning models must be smoothly incorporated into the clinical workflow in order to be used for diabetes prediction. As depicted in figure.2, the integration process entails giving careful consideration to data collection, preprocessing, model training, result communication, and decision support. This section focuses on the process for incorporating the clinical workflow for diabetes prediction with the machine learning model.

Firstly, patient data is collected, which includes relevant medical records, laboratory results, and patient demographics. This data is then pre-processed to handle missing values, outliers, and standardize the variables. Feature engineering techniques are applied to extract meaningful features from the data, such as glucose levels, body mass index, and medical history. Subsequently, the pre-processed data is utilized

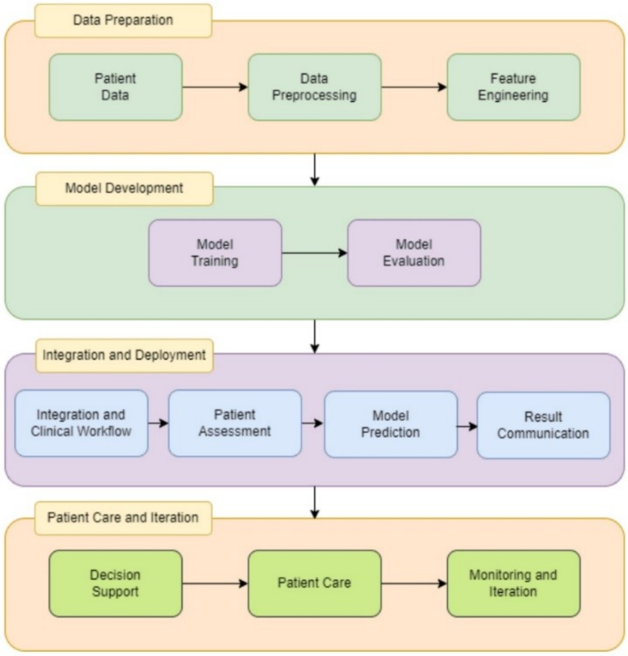

Figure.2: Integration of Machine Learning Model into the Diabetes Clinical Workflow

to train machine learning models. For diabetes prediction, diverse algorithms like support vector machines, logistic regression, and deep learning models can be employed. The trained models are evaluated to assess their performance and determine their accuracy, sensitivity, specificity, and other relevant metrics.

Once the models are validated, they are integrated into the clinical workflow. The models receive patient assessments and generate predictions regarding the likelihood of diabetes. These predictions are communicated to healthcare providers and patients, enabling informed decision-making and personalized care planning.

The model's results serve as decision support for clinicians, aiding in the determination of appropriate treatment plans, lifestyle modifications, or the need for further diagnostic tests. Additionally, the patient's assessment, along with the model's predictions, guides the communication of the results to the patient, fostering shared decision-making and patient engagement. To ensure the continuous monitoring and improvement of the model's performance, regular monitoring and iteration are essential. This includes tracking the model's predictions, evaluating its accuracy over time, and updating the model based on new patient data and feedback.

The integrated clinical workflow for diabetes prediction faces challenges that need to be addressed. Some of the factors that need to be considered include the interpretability and explain ability of machine learning models, the smooth integration of these models into existing clinical workflows, and the ethical aspects related to data privacy and security.

IV. Literature Review

In recent years, machine learning has become a pivotal innovation in medicine, with a promising outlook for the future. This study aims to employ machine learning classifiers to categorize diabetes patients based on their self-reported information and clinical conditions. We provide an overview of research conducted over the past decade to identify shortcomings in existing works related to machine learning classifiers for diabetes treatment strategies.

Sun and Zhang [26] investigated a range of deep learning and classification techniques, such as support vector machines, decision trees, random forests, and artificial neural networks. The authors [27] used a logistic regression-based classification technique to classify diabetes-related data in different research. There were 459 patients in the training dataset and 128 patients in the testing dataset. The logistic regression model is noteworthy for achieving a high 92% classification accuracy. It's important to note, nevertheless, that this model's validation was limited because it wasn't compared to other diabetes prediction models that are currently in use. For training and testing, the dataset was divided in half.

Naive Bayes, decision trees, and SVM learning techniques were used by researchers in their examination of the Pima Indians Diabetes Collection [28]. Notably, when it came to predicting diabetes, the Naive Bayes classifier showed the best accuracy. Using 10 equal pieces of the dataset—nine for training and one for assessment—Sisodia used tenfold cross-validation. Precision, accuracy, recall, and area under the curve were employed in the assessment as conventional evaluation criteria for diabetes prediction.

The authors in [29] evaluated a number of machine learning approaches in their study. In particular, they assessed how well neural networks (NN), random forests, and Naive Bayes performed. The Matthews correlation coefficient was the assessment metric used by the authors to determine how effective these strategies were.

To extract relevant features from the Pima Indians Diabetes Dataset, the authors in [30] used two different feature selection techniques: Principal Component Analysis (PCA) and Linear Discriminant Evaluation. Factor analysis approaches include Principal Component Analysis and Linear Discriminant Analysis. A comparative comparison of attribute selection procedures was also included in the study. The authors used the dataset under examination to test a number of machine learning techniques, such as the AdaBoost, K-Nearest Neighbors (KNN), and Radial Basis Kernel, for the classification job.

A single diagnosis strategy for early-stage diabetes is clearly not very successful, as demonstrated by the Gujral Writing Survey of Diabetes Assumptions findings [31]. Artificial Neural Networks (ANN) incorporate numerous classifiers, such Evolutionary Algorithms, Principal Component.

Analysis, and Support Vector Machines (SVM), to get the best results.

In the study of the Pima Indians Diabetes Dataset, the researchers[32] made noteworthy contributions. The importance of variables including BMI, blood glucose levels, and the number of pregnancies in the dataset was highlighted by their analysis. Using logistic regression and RStudio, they estimated accuracy and obtained an accuracy rate of 75.32%.

In their examination of the Pima Indians Diabetes Dataset, the authors in [33] used a variety of models, such as the multilayer perceptron, Bayes net, Hoeffding tree,and JRip. Both the greedy iterative and the best initial feature selection techniques were used in the research to increase classifier effectiveness. The researchers chose just four characteristics from the complete eight: age, BMI, diabetes pedigree function, and plasma glucose level. With a recall score of 76.2% and an accuracy rate of 75.7%, the Hoeffding tree approach in particular showed remarkable results.

Diabetic complications can be rather serious and the illness spreads quickly. Though trustworthy statistics are hard to come by, early diagnosis lowers risks. With the help of feature selection, hyperparameter optimization, and missing value imputation, the authors in [34] can provide a weighted ensemble of machine learning classifiers (NB, RF, DT, XGB, and LGB) and introduce a SA dataset. The new dataset for reliable diabetes prediction models employing population-level data is beneficial to the ensemble (DT + RF + XGB + LGB) with our preprocessing, as demonstrated by the considerable improvement in prediction (0.735 accuracy and 0.832 AUC).

Support Vector Machine (SVM) and Artificial Neural Network (ANN) models are employed in a fused machine learning (ML) technique presented in another study in [35]. The final diabetes diagnosis made by the fuzzy logic system is based on real-time medical information, and the dataset is split into training and testing data in a 70:30 ratio. With an astounding accuracy of 94.87%, the suggested fused model outperforms earlier approaches.

A low-code Pycaret machine learning approach is used in another study [36] for the categorization, detection, and prediction of diabetes. Gradient boosting emerges as the most accurate when many classifiers are hyper-tuned; it achieves 90% accuracy, outperforming other machine learning classifiers.

V. Discussion and Challenges

A range of machine learning techniques for diabetes prediction are presented in the evaluated literature. The following are significant flaws and difficulties, which include the requirement for larger and more varied datasets, the investigation of deep learning methodologies, and the development of strict model comparison procedures.

- The literature frequently mentions the drawback of diabetes prediction based on a limited set of characteristics. The predictive power of publicly accessible datasets, like the Pima Indians Diabetes Collection, may be limited since they sometimes only include a small number of variables. This emphasizes that in order to improve forecast accuracy, feature engineering or the inclusion of additional data sources are required.
- Some research exclude data that isn't full, which reduces the size of the dataset and could affect how reliable the findings are. Robust handling of missing data is necessary to guarantee the accuracy of forecasts.
- The evaluated research has not made full use of deep learning techniques like recurrent neural networks. More precise and effective diabetes prediction systems may be produced by investigating sophisticated deep learning algorithms.
- Many research restricts the validation and benchmarking of their techniques by failing to compare their suggested models with the current diabetes prediction models. Analytical comparisons may shed light on how well certain methods perform in relation to one another.

VI. Conclusion

The review article has extensively examined the use of machine learning for diabetes prediction, offering a detailed analysis of its applications. The review highlighted various machine learning algorithms, methodologies, and datasets used in previous studies, along with their contributions to the field. The integration of machine learning models into the clinical workflow has shown promising results in improving the prediction of diabetes and its related complications. These models have demonstrated their effectiveness in risk stratification, early detection, and personalized interventions, leading to better patient outcomes and management of the disease.

However, several challenges need to be addressed for the widespread adoption of machine learning-based diabetes prediction models. These include Interpretability and Explainability of the models, seamless integration into clinical workflows, ethical considerations regarding data privacy and informed consent, and ensuring generalizability and external validation across diverse populations. Future directions in this field involve the development of robust and interpretable models, exploring novel data sources and features, assessing long-term outcomes and clinical utility, and promoting personalized risk assessment and continuous monitoring.

References

[1] International Diabetes Federation https://idf.org fetched 20.09.2023

[2] Pitt, Bertram, et al. "Cardiovascular events with finerenone in kidney disease and type 2 diabetes." New England journal of medicine 385.24 (2021): 2252-2263.

[3] Jyotismita Chaki, S. Thillai Ganesh, S.K Cidham, S. Ananda Theertan, Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 6, Part B, 2022, Pages 3204-3225, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2020.06.013.

[4] Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. J Diabetes Metab Disord. 2020 Apr 14;19(1):391-403. doi: 10.1007/s40200-020-00520-5.

[5] LaValley, Michael P. "Logistic regression." Circulation 117.18 (2008): 2395-2399

[6] Al Jarullah, Asma A. "Decision tree discovery for the diagnosis of type II diabetes." 2011 International conference on innovations in information technology. IEEE, 2011.

[7] N. Mohan and V. Jain, "Performance Analysis of Support Vector Machine in Diabetes Prediction," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1-3, doi: 10.1109/ICECA49313.2020.9297411.

[8] El_Jerjawi, Nesreen Samer, and Samy S. Abu-Naser. "Diabetes prediction using artificial neural network." (2018).

[9] Aslan MF, Sabanci K. A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data. Diagnostics (Basel). 2023 Feb 20;13(4):796. doi: 10.3390/diagnostics13040796. PMID: 36832284; PMCID: PMC9955314.

[10] Priyanka Rajendra, Shahram Latifi, Prediction of diabetes using logistic regression and ensemble techniques, Computer Methods and Pro-

grams in Biomedicine Update, Volume 1, 2021, 100032, ISSN 2666-9900, https://doi.org/10.1016/j.cmpbup.2021.100032.

[11] W. Chen, S. Chen, H. Zhang and T. Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2017, pp. 386-390, doi: 10.1109/ICSESS.2017.8342938.

[12] K. VijiyaKumar, B. Lavanya, I. Nirmala and S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878802.

[13] Patil, Ratna, et al. "A modified mayfly-SVM approach for early detection of type 2 diabetes mellitus." Int. J. Electr. Comput. Eng 12.1 (2022): 524.

[14] M. T. Islam, H. R. H. Al-Absi, E. A. Ruagh and T. Alam, "DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only," in IEEE Access, vol. 9, pp. 15686-15695, 2021, doi: 10.1109/ACCESS.2021.3052477.

[15] Farajollahi, Boshra, et al. "Diabetes diagnosis using machine learning." Frontiers in Health Informatics 10.1 (2021): 65.

[16] Rufo, Derara Duba, et al. "Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM)." Diagnostics 11.9 (2021): 1714.

[17] Ashima Singh, Arwinder Dhillon, Neeraj Kumar, M. Shamim Hossain, Ghulam Muhammad, and Manoj Kumar. 2021. EDiaPredict: An Ensemble-based Framework for Diabetes Prediction. ACM Trans. Multimedia Comput. Commun. Appl. 17, 2s, Article 66 (June 2021), 26 pages. https://doi.org/10.1145/3415155

[18] Taha, Altyeb Altaher, and Sharaf Jameel Malebary. "A hybrid meta-classifier of fuzzy clustering and logistic regression for diabetes prediction." Computers, Materials and Continua 71.2 (2022): 6089-105.

[19] A. Yahyaoui, A. Jamil, J. Rasheed and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," 2019 1st International Informatics and Software Engineering Conference (UBMYK), Ankara, Turkey, 2019, pp. 1-4, doi: 10.1109/UBMYK48245.2019.8965556.

[20] Arora, Nitin, et al. "A Novel Architecture for Diabetes Patients' Prediction Using K-Means Clustering and SVM." Mathematical Problems in Engineering 2022 (2022).

[21] Jader, Rasool, and Sadegh Aminifar. "Fast and Accurate Artificial Neural Network Model for Diabetes Recogni-tion." NeuroQuantology 20.10 (2022): 2187-2196.

[22] Yadav, D.C., Pal, S. (2021). An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Disease. In: Mathur, R., Gupta, C. P., Katewa, V., Jat, D. S., Yadav, N. (eds) Emerging Trends in Data Driven Computing and Communications. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore. https://doi.org/10.1007/978-981- 16-3915-9_18

[23] Kopitar, L., Kocbek, P., Cilar, L. et al. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep 10, 11981 (2020). https://doi.org/10.1038/s41598-020-68771-z

[24] Tuppad, Ashwini, and Shantala Devi Patil. "Machine learning for diabetes clinical decision support: a review." Advances in Computational Intelligence 2.2 (2022): 22.

[25] Berchialla, Paola, et al. "Prediction of treatment outcome in clinical trials under a personalized medicine perspective." Scientific Reports 12.1 (2022): 4115.

[26] Sun, Y. L. & Zhang, D. L., 2019. Machine Learning Techniques for Screening and Diagnosis of Diabetes: A Survey. Technical Gazette, Volume 26, p. 872– 880

[27] Qawqzeh, Y. K. et al., 2020. Classification of diabetes using photoplethysmogram (PPG) waveform analysis: logistic regression modeling. BioMed Research International, pp. 1-20.

[28] Sisodia, D. & Sisodia, D. S., 2018. Prediction of diabetes using classification algorithms. Procedia Computer Science, Volume 132, p. 1578–1585.

[29] Hussain, A. & Naaz, S., 2021. Prediction of diabetes mellitus: comparative study of various machine learning models. Advances in Intelligent Systems and Computing, Volume 1166, p. 103–115.

[30] Choubey, D. K. et al., 2020. Comparative analysis of classification methods with PCA and LDA for diabetes. Current Diabetes Reviews, 16(8), p. 833–850.

[31] Gujral, S., 2017. Early diabetes detection using machine learning: a review. International Journal for Innovative Research in Science & Technology, 3(10), pp. 45-60.

[32] Tigga, N. P. & Garg, S., 2020. Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, Volume 167, p. 706–716.

[33] Mercaldo, F., Nardone, V. & Santone, A., 2017. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Computer Science, Volume 112, p. 2519–2528.

[34] Dutta A, Hasan MK, Ahmad M, Awal MA, Islam MA, Masud M, Meshref H. Early Prediction of Diabetes Using an Ensemble of Machine Learning Models. *International Journal of Environmental Research and Public Health*. 2022; 19(19):12378. https://doi.org/10.3390/ijerph191912378

[35] U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, doi: 10.1109/ACCESS.2022.3142097.

[36] Whig, P., Gupta, K., Jiwani, N. et al. A novel method for diabetes classification and prediction with Pycaret. Microsyst Technol 29, 1479–1487 (2023). https://doi.org/10.1007/s00542-023-054732