International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.phpa

A CASE STUDY OF MACHINE LEARNING METHODOLOGIES FOR THE DIABETIC PREDICTION

[1]Mr.K. Sathiesh Kumar,
PhD Research Scholar, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Coimbatore

[2]Dr.K. Geetha,
Associate Professor, Department of Computer Application, Sri Krishna Adithya College of Arts and Science, Coimbatore

Abstract Early accurate prediction technologies are essential for managing long-term diabetes because the disease remains a significant global health issue. Metabolic disorder known as diabetes causes high blood sugar in individuals because the body fails to produce or use insulin effectively. Uncontrolled diabetes results in health complications that affect the heart as well as cause damage to kidneys and nerves. This research inquiry focuses on different Machine Learning (ML) methods used to predict diabetes risks as well as the identification of its symptoms. Multiple classification technologies such as Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Logistic Regression (LR) and k-Nearest Neighbors are assessed which successfully detect diabetic patterns. The study analyzes data pre-processing techniques which include feature normalization as well as outlier treatment and class sampling because these influence model accuracy levels. The survey highlights the evaluation methods through cross-validation and utilizes accuracy, precision, recall, F1-score and ROC curves to maintain high prediction reliability and robust prediction outcomes. The assessment conducts the thorough analysis of critical difficulties related to data privacy together with model interpretability problems and techniques for managing imbalanced datasets. This survey investigates the combination of medical records with sensor-based data through secure data-sharing techniques for better prediction system performance. Researchers along with healthcare practitioners have the required comprehensive base to understand ML solutions for diabetes prediction through the survey data.
Keywords: Accuracy, Decision Trees (DT), Diabetes, Machine Learning (ML), Prediction, Support Vector Machines (SVM)

INTRODUCTION:

The prediction of diabetes depends on recognizing major medical and lifestyle indicators which heighten the disease susceptibility. Predictive models such as clinical assessments make use of these factors as the essential elements for input. The main reason for diabetes development stems from high Body Mass Index (BMI). Yet, severe diabetes-related health risks emerge when fat accumulates primarily in abdominal regions. Type 2 diabetes develops frequently in patients with obesity since the condition blocks insulin sensitivity. Family history stands as one of the essential risk factors for the development of the disease. Diabetics show higher probabilities for passing the disease to close relatives because of inherited genetic factors. The risk of Type 2 diabetes increases after age 40 until going beyond this threshold. Thereby, vulnerability is extended to the condition. The condition of blood pressure presents an important sign because the individuals suffering from diabetes typically experience high blood pressure levels which aggravate insulin resistance. Direct symptoms of future diabetes include both elevated fasting blood sugar levels and unexpected results through glucose tolerance testing. Higher cholesterol and triglycerides present in the body increase patient risk since these indicate problematic metabolic processes. It needs the assessment of both active and passive lifestyles since these elements strongly impact the efficient management of glucose levels. Processing food consumption, high sugar drink intake and minimal fiber intake from diets function as the major lifestyle causes which boost diabetes prediction. Healthcare providers check gestational diabetes history together with PCOS among female patients when assessing possible diabetes risks. Healthcare professionals use cause monitoring to screen people with the potential risks of developing diabetes so that preventive measures are taken for delaying or preventing disease onset. The chronic metabolic condition which is diabetes creates widespread body effects when the individuals fail to control the disease effectively. The health risks stemming from diabetes emerge in both right away and long run as well as include the various problems starting from standard problems that escalate to dangerous health situations. An immediate result of diabetes consists of impaired glucose metabolism. The inability to manage insulin properly leads to high bloodstream glucose levels known as hyperglycemia when insulin production falls short of the needed amount. The continuous presence of elevated blood sugar levels eventually damage blood vessels as well as fundamental organs. Cardiovascular disease stands as one of the primary serious conditions that affect the diabetes patients. High blood sugar levels cause vessel lining damage which increases the chances of heart attacks, strokes as well as hypertension development. The painful and weakening effects along with numbness of hands and feet occur from nerve damage triggered by diabetes which is known as diabetic neuropathy. The combination of diabetes with foot ulcers leads to severe healthcare challenges that eventually result in patients needing foot amputations. Diabetic nephropathy develops when blood sugar causes stress to kidney filtration resulting in kidney failure. Diabetic retinopathy occurs because of diabetes and damages the eyes which affects vision and eventually leads to blindness. Having weak immunity is a result of diabetic neuropathy since it leaves the body vulnerable to infections and prevents proper wound healing.Multiple causes exist which lead to diabetes formation. The most relevant cause comes from genetic background. The probability of becoming diabetic boosts significantly when a person possesses diabetes within the immediate family members. A diet which consists of excessive sugar, fats and processed foods with an unsatisfactory lifestyle pattern plays an essential role in diabetes development. The body's glucose level regulation becomes ineffective when the people do not exercise enough or spend too much time sitting down.Weight gain especially abdominal fat build-up leads to insulin resistance which strongly promotes the development of Type 2 diabetes. The adult population experiences higher diabetes risks which further increases based on age advancement. South Asian people along with Native Americans and African descent are more likely to develop diabetes than other ethnic groups. The same is true for people of Hispanic descent. Multiple health conditions comprising hypertension, high cholesterol and Polycystic Ovary Syndrome (PCOS) as well as the previous experience of gestational diabetes enhance risk levels. Specific social conditions like healthcare shortages and poor health education combined with economic disparities produce additional conditions that increase diabetes prevalence.**Motivation:** The increasing number of diabetes cases in the world produces health issues so early detection serves as the vital step for both disease prevention and effective management. The large collection of medical data enables ML to detect patterns and anticipate diabetes susceptibility through its methodologies. The predictive methods deliver legitimate along with prompt and accurate outcomes that assist healthcare providers to deliver the appropriate treatment options. The research aims to examine ML techniques analyze diabetic diagnosis potential while assessing the influence on healthcare results.

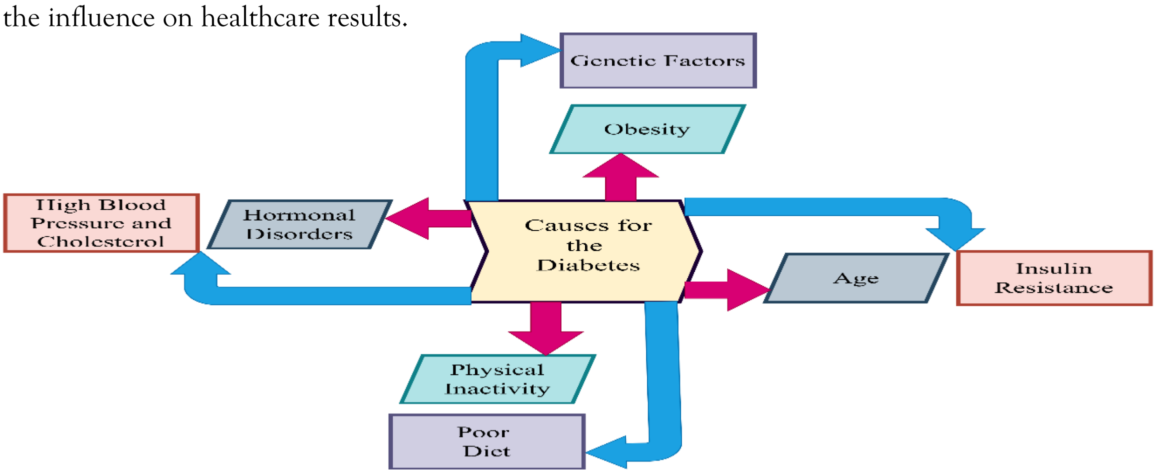

Figure 1: Vital Reasons for the Diabetes

Figure 1 displays the insufficient reasons behind Diabetes onset. The disease has multiple underlying reasons. Examples of causative factors include being obese and inactive combined with aging and low diet levels as well the genetical factors, high cholesterol and blood pressure rates. The combination of overweight condition and sedentary lifestyle together with poor eating patterns, advancing age and genetic predisposition causes Diabetes development. The disease leads to severe health complications which affect the heart, nerves, damage kidneys, eyes as well as triggering emotional stress. Additional health risks for diabetes development originate from high blood pressure as well as high cholesterol levels together with PCOS. Timely medical checking combined with appropriate healthcare maintains or eliminates diabetes development.

LITERATURE REVIEW

Researchers use various studies which explore ML detection methods for diabetes yet analyze both medical records and life-style factors. The integration of Decision Trees (DT) with Support Vector Machines (SVM) along with Logistic Regression (LR) develops the outstanding abilities for pre-diabetic symptom detection. Researchers need appropriate feature selection methods combined with high-quality collected data in order to achieve the results. Medical professionals use these available methods to track patients requiring quick medical assistance.The research conducted by Ismail and Materwala [21] developed a diabetes prediction intelligent system that implemented various ML algorithms. These diagnosis outcomes improved due to the automatic learning possibilities and data analysis features. In medical practice, the research-based tools assisted healthcare providers to make decisions about patient care for individual treatment needs. The research investigation produced effective results for determining medical patient susceptibility to diabetes.The PIMA Indian data served as a basis for Salih et al. [22] when these authors constructed ML models for diagnostic purposes. Multiple algorithms were evaluated by the authors to assess the accuracy levels when processing genuine clinical medical data. Experienced researchers highlighted in the research that the combined effect of data handling and preprocessing techniques led to better model reliability. ML detection methods demonstrated the effectiveness for diagnosing diabetes at its beginning stages.Natarajan et al. [23] established a diabetes prediction ensemble method that had not selected any specific modeling approach. The method used various elements to increase the accuracy levels of various classifiers. The technique managed to determine vital health metrics associated with diabetes.

Table 1: Comparison table on the objectives presented in the various diabetes Prediction works

Authors	Technique Used	Dataset/ Approach	Objective	Key Outcome
Hasan et al. [24]	Stacked Ensemble Approach	Empirical modeling with ensemble learning	To predict diabetes-positive cases more accurately	Achieved improved prediction accuracy by combining multiple classifiers
Alnowaiser [25]	KNN Imputation + Tri-Ensemble Model	Feature-imputed clinical dataset	To enhance prediction in diabetic patients using imputed features	Outperformed traditional models by integrating multiple learners with KNN-based preprocessing
Jose et al. [26]	ML Models + Intervention System	Cardiovascular health data in diabetics	To manage cardiovascular health in diabetic patients	Enabled precise prediction and preventive healthcare interventions

Teki et al. [27]	Mean Shift Clustering	Unsupervised clustering on health data	To create a prediction system for diabetes using clustering	Provided early detection through cluster-based analysis
Chatterjee et al. [28]	Feedforward Neural Network (FNN) + Oppositional Whale Optimization Algorithm	Optimized feature selection from health records	To enhance prediction accuracy using hybrid optimization and Deep Learning (DL)	Delivered high performance through optimal feature tuning and DL integration

The studies analyzing different ML solutions for diabetes forecasting and management appear in Table 1. The research applied stacked ensemble modeling to reach better accuracy outcomes through classifier combination and obtained higher results by integrating KNN-imputed features within tri-ensemble modeling. The systems are developed using prediction methods for diabetic patients by delivering individualized medical services. Two different approaches were implemented for mean shift detection clustering for early warning signals but neural networks as well as metaheuristic algorithms are utilized for optimization. Krishnamoorthi et al. [29] introduced a framework based on various MLtechniques for diabetes prediction. Intelligent prediction models were the main target of this research to improve healthcare support capabilities. The research presented initial evidence about better diagnostic measurements. The study maintains minimal impact because researchers doubt its measurement methods. Sheng et al. [30] investigated the value of non-traditional along with traditional lipid tests to detect future diabetes risks in non-diabetic participants. Research relied on medical testing techniques to prove connections between diabetes occurrence and lipid measurements. The research verified that non-standard lipid assessment markers enable to achieve superior risk assessment abilities. The researchers from Lama et al. [31] developed a prediction system for diabetes risk evaluation that analyzed middle-aged Swedish individuals. Predictive models developed the knowledge about patient risk distribution through the analysis of health data sources and population demographic records. Forecast accuracy turned out to be exceptional since the prediction models incorporated both lifestyle data and physiological measurement results. The method allowed health programs to execute early preventive actions within the programs. Rajesh et al. [32] created statistical analysis to find various gene patterns linked to diabetic macular edema pathogenesis. Through this approach, researchers managed to enhance genetic risk factor interpretation while creating better predictive models. The study created tactics to identify individual patient complications arising from diabetes.

Table 2: A precise comparison table related to diabetes studies

Study	**Focus Area**	**Input Features**	**Outcome Predicted**	**Statistical Techniques**
Zhang et al. [33]	Predicting visual acuity after anti-VEGF therapy in DME	OCT features, baseline BCVA, treatment duration	Post-treatment visual acuity	Ensemble regression techniques
Li et al. [34]	Detecting DME through clinical correlation graphs	Demographics, lab tests, eye exam findings	Diagnosis of DME	Knowledge graph relationships

Fan et al. [35]	Predicting diabetic complications & poor control	Medication adherence, HbA1c, BMI, comorbidities	Risk of complications & glycemic instability	Logistic regression, classifiers
Liu et al. [36]	Predicting treatment outcome in DME patients	Baseline vision, injection history, retinal data	BCVA improvement post-treatment	Combined decision models
Sumathy et al. [37]	Diabetic retinopathy diagnosis from EHR	Fasting glucose, blood pressure, age, vision status	DR stage diagnosis	Tree-based classification models
Cao et al. [38]	Predicting response to anti-VEGF from OCT	Retinal thickness, layer disruption, fluid accumulation	Treatment response category	Imaging-based learning, feature extraction

The analysis of six research studies focused on diabetic eye complication forecasting appears in table 2. Each research in this table demonstrates its medical purpose in addition to showing how features were analyzed and predictive results were obtained through statistical approaches. The study evaluates how patients receive their disease diagnosis and how their treatment will affect their health condition. As an efficient tool the table provides an effective overview of data-based techniques used in diabetic medical care. PPARG gene variations were associated with diabetic risks form the research topic of Mustafa et al. [39]. This screening method identified specific nucleotide base modifications that resulted in structural proteins dysfunction. Digital computer systems conducted biological effect predictions for each genetic change discovered. This experiment demonstrated that certain changes in genetic sequences produced diabetes symptoms that affect protein activity. The research article by Soni and Varma [40] examined statistical techniques to enhance diabetes risk procedures. The assessment used clinical indicators that measured glucose blood pressure, BMI and age to determine patient status. Research executed multiple confirmation tests to generate a model that enabled the most efficient and earlier detection of diabetes. This research showed that identifying vulnerable individuals strongly depended on organized clinical medical information. he research by Dutta et al. [41] centered its investigation on healthcare markers related to the patients for diagnosing diabetes onset. The evaluation included both insulin measurements and glucose test results in addition to the whole body measurements analysis. Research results from statistical models produced better predictive values. Research at the prevention stage had improved through early diagnosis identification because the study revealed future patients with the condition.

Table 3: Various methodological comparisons utilized in the prediction of diabetes

Authors	**Methodology Focus**	**Focus**	**Algorithm/ Technique Used**	**Key Findings**
Ahmed et al. [42]	Web Development for Health	Diabetes prediction using a smart web platform	LR, SVM	Developed a smart web application for diabetes prediction using ML
Shin et al. [43]	Model Comparison and Performance	Analysis of different diabetes prediction models	DT, SVM, Convolutional Network (CNN)	Developed and compared various ML models for diabetes prediction

Zhu et al. [44]	Personalized Medicine and Artificial Intelligence (AI)	Blood glucose prediction for Type 1 Diabetes	Evidential DL, Meta-Learning	Proposed a personalized diabetes prediction model using DL
Zhu et al. [45]	Edge Computing and DL	Population-specific glucose prediction	Transformer-based DL	Introduced a transformer-based model for glucose prediction in diabetes care
Daliya & Ramesh [46]	Cloud-Based Risk Prediction Model	Risk prediction for diabetic progression	Ensemble Methods, Azure ML	Developed an optimized ensemble model for diabetes progression risk prediction

Different strategies for predicting diabetes along with disease progression risk evaluation are examined using Table 3 as a comparison. Web solutions for diabetes prediction form part of the research domain but the studies are about model assessments and DL analysis of blood glucose predictions. The combination of edge computing with ensemble methods operates in an advanced system to boost blood glucose prediction precision and diabetes advancement risk evaluation capabilities. The research by Shams et al. [47] presented a novel RFE-GRU model for classifying diabetes patients using PIMA Indian dataset information. This method achieved effective feature selection for diabetes prediction applications after the implementation as a part of Recursive Feature Elimination (RFE) for the study analysis. The authors used this model to enhance identification accuracy of diabetes patients. The research study showed that RFE-GRU exhibited strong potential to achieve better predictions within medical information platforms. The authors from [48] created an optimized XGBoost model that used Bayesian optimization for optimizing its hyperparameters to forecast diabetes cases. XGBoost analysis led the research team to study this ML algorithm because it was popular in practice while these optimized its parameters through Bayesian optimization for better prediction outcomes. The proposed method used a new diabetes diagnosis approach that outperformed the traditional diagnostic strategies. This research showed optimization methods improving medical model functions specifically intended for healthcare use. Using causal discovery and inference models as well as ML techniques, Noh and Kim [49] predicted diabetes. The researchers used causal linkages between the elements in ML models to increase the accuracy of diabetes start estimates. The results indicate that causal discovery had assisted to increase the dependability and comprehensibility of prediction models. Based on the evidence, this combined approach improved the prediction of complicated diseases including diabetes by targeting underlying cause elements. Sajid et al. [50] created a hybrid ensemble model to identify diabetes and offered individualized dietary and physical activity recommendations. The researchers developed a consistent model for diabetes diagnosis and treatment plan suggestions using many ML techniques. This new method aimed to improve diagnosis accuracy and gave individual advice to the patients. According to the study, in hospital environments especially for managing chronic conditions like diabetes, ensemble learning was especially useful. Among the machine learning methods that have been accurate in several studies trying to forecast diabetes using medical and lifestyle data are LR, DT and SVM. These techniques underline the need of feature selection and datasets in generating reliable predictions. Therefore, it assisted in the earlier diagnosis and control of diabetes.

PROBLEM IDENTIFICATION

A chronic disorder, Diabetes causes major problems if not under control and affects millions of people globally. Conventional methods of diagnosis which mostly rely on hand inspection have a tendency to overlook tiny signs. Accurate, fast as well as scalable predictive algorithms that call for physiological and clinical data have to be developed. Modern medical facilities find it difficult to properly handle the enormous volumes of patient records. Early diabetes are predicted and better judgments made identify the tendencies in this data.

Surveyed Methodologies

In this survey, several papers are examined based on the prediction of Diabetes. The following contents explain about some of the analyzed methodologies.

Machine Learning-based Diabetic Prediction

The worldwide untreated diabetes epidemic cause major issues. Early detection lowers long-term health risks and lets quick medical action possible. Machine learning study of patient data is expected to expose the unidentified trends and risk factors. Deep learning algorithms considerably raise diabetes prediction accuracy, according to the research using datasets like PIMA Indian dataset [3]. Furthermore, explainable ML technology has made medical practitioners' prediction process more consistent and intelligible [4]. These developments point to a direction towards more evidence-based and efficient diabetic treatment.

Combination of Deep Learning and Feature Selection Strategies

Combining CNNs and Bidirectional Long Short-Term Memory (Bi-LSTM) networks [13] enhance the prediction model. This hybrid architecture understands the complicated spatial properties by means of CNNs to identify the sequential relationships and Bi-LSTMs. Its aim is to manage data in real time. The method shines in hectic hospital environments where the accurate projections are absolutely vital. Instead of depending just on stationary characteristics, an adaptive ensemble technique [23] is offered for data processing. The model-agnostic approach guarantees that only pertinent variables are assessed during the prediction process by carefully gathering the most significant features from vast clinical datasets. By reducing computational complexity, it not only improves the accuracy of the model but also makes it more flexible and compatible to other ML approaches. Notwithstanding these differences, all studies have as the objective early stage diabetes detection and diagnosis utilizing intelligent data analysis. These show that the combination of DL with feature selection enhance the decision-making capacity of medical practitioners. Hence, the patient outcomes are improved.

Collaboration of the Machine Learning and Genetical Strategies

Machine learning technologies are under great investigation to increase the accuracy and efficiency since the usefulness in diabetes prediction drives the demand. One approach is to merge Feedforward Neural Networks (FNN) and optimizing strategies including the Oppositional Whale Optimisation Algorithm. Diabetic data is classified [28] using FNN and optimisation method to modify model parameters. Hence, the efficiency of this hybrid model is established and the predicting capability is improved. Moreover, [39] focused especially on computationally searching the PPARG gene SNPs associated to diabetes. By examining the effect of PPARG gene mutations on protein activities and contribution to diabetic outcomes, this work showed that the genetic elements are absolutely essential for diabetes prediction. The former rapidly applied ML to clinical data to generate real-time predictions whereas the latter offered insights into genetic predispositions that affected prediction models. Apart from providing some overview of the several methods of diabetes prediction, these results highlight the possibilities of genetic analysis and ML to enhance the earlier identification and patient outcomes.

Methodological Assessment of Diabetes

Increasing numbers of researches using ML techniques improve the treatment and prediction of diabetes. Using Azure ML to improve the accuracy and adaptability of real-time healthcare applications, an optimal ensemble model [46] is suggested for cloud-based diabetes progression risk forecasting. This model combines several ML approaches. This method increases the scalability and performance of diabetes course prediction by using cloud computing. Based on the PIMA Indian dataset, a new RFE-GRU model is developed [47] for diabetes classification. Focused on the most significant features, these used the combination of RFE and Gated Recurrent Units (GRU) to lower the computational cost and raise classification accuracy. Both studies show that ML improves diabetes prediction even if these apply different approaches. This is achieved by using DL methods to raise classification accuracy or by using cloud computing to generate real-time risk prediction. At last, this could result in more prompt and successful medical interventions. The several ML techniques for diabetes incidence prediction are investigated in this work. Emphasizing the mix of genetic analysis, feature selection techniques and DL, it shows that these approaches increase prediction accuracy, lower computational complexity and enable early diagnosis. Thereby, more efficient and timely healthcare actions are guided.

DISCUSSION

The approaches investigated in this survey clearly show that ML, especially DL models and advanced feature selection techniques is finding growing uses in diabetes prediction. Especially DL, ML approaches have shown great accuracy for diabetes prediction, surpassing more traditional approaches. Early diagnosis depends on DL algorithms being able to identify latent trends and patterns in complex as well as vast clinical datasets. Moreover, explainable ML techniques guarantee that forecasts are rational. When used with feature selection methods including ensemble approaches and optimization algorithms, DL lowers computer overhead, adapts to different healthcare environments and removes duplicate or superfluous data features. Thereby, prediction accuracy is improved.

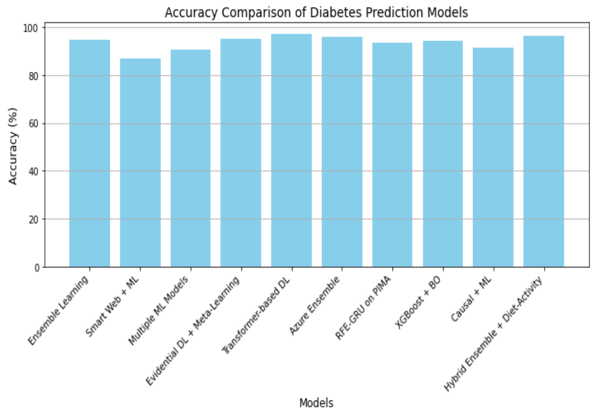

Figure 2: Accuracy Comparison Graphs on Diabetes Forecasts Models

Figure 2 presents the accuracy score comparisons for numerous diabetes prediction methods. The graph above displays the accuracy values ranging from 80 to 100. The comparison values vary slightly. Moreover, the combination of ML and genetic data reveals the greater possibility for the creation of diabetes prediction systems. By including the genetic data such as variations in the PPARG gene into ML algorithms, researchers are creating models that more fairly explain a person's diabetes status. Thus, these produce more precise forecasts and modify therapy to fit particular patients. In fast changing healthcare systems, the scalability and agility of the cloud have also been rather important in allowing the real-time diabetes risk prediction. Diabetes prediction and management are more accurate and efficient by aggregating cloud-based technology, genetic analysis, feature selection and DL. Better patient outcomes and more effective healthcare operations depend on these developments greatly improving early identification and risk prediction.

CONCLUSION

In conclusion, ML seems quite accurate for raising the early stage diabetes screening accuracy and efficiency. Evaluating intricate medical data and spotting important risk factors calls for DL, feature selection and cloud computing among other beneficial techniques. Combining the innovative technologies such as Bi-LSTM, CNN and RFE assists the researchers to minimize computer complexity while increasing prediction accuracy. Including genetic data allows further customizing prediction models and provides a significant layer of understanding on the diabetes risk. According to the investigated studies, multidisciplinary methods of diabetes control are useful and successful. The precise and timely projections produced by ML advancements have great potential to enhance healthcare operations. Therefore, this aids to raise the effectiveness of medical therapies and improve patient outcomes.

REFERENCES

1. Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access*, *8*, 76516-76531.
2. Rani, K. J. (2020). Diabetes prediction using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *6*(4), 294-305.
3. Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*, 391-403.
4. Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. *Healthcare technology letters*, *10*(1-2), 1-10.
5. Rajendra, P., & Latifi, S. (2021). Prediction of diabetes using logistic regression and ensemble techniques. *Computer Methods and Programs in Biomedicine Update*, *1*, 100032.
6. Zhou, H., Myrzashova, R., & Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. *EURASIP Journal on Wireless Communications and Networking*, *2020*, 1-13.
7. Bukhari, M. M., Alkhamees, B. F., Hussain, S., Gumaei, A., Assiri, A., & Ullah, S. S. (2021). An improved artificial neural network model for effective diabetes prediction. *Complexity*, *2021*(1), 5525271.
8. Ahmed, U., Issa, G. F., Khan, M. A., Aftab, S., Khan, M. F., Said, R. A., ... & Ahmad, M. (2022). Prediction of diabetes empowered with fused machine learning. *IEEE Access*, *10*, 8529-8538.
9. Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked*, *17*, 100179.
10. Olisah, C. C., Smith, L., & Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*, *220*, 106773.
11. Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., ... & Louni, A. (2022). A classification algorithm-based hybrid diabetes prediction model. *Frontiers in Public Health*, *10*, 829519.
12. Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, *2*, 40-46.
13. Madan, P., Singh, V., Chaudhari, V., Albagory, Y., Dumka, A., Singh, R., ... & AlGhamdi, A. S. (2022). An optimization-based diabetes prediction model using CNN and Bi-directional LSTM in real-time environment. *Applied Sciences*, *12*(8), 3989.
14. Dai, L., Sheng, B., Chen, T., Wu, Q., Liu, R., Cai, C., ... & Jia, W. (2024). A deep learning system for predicting time to progression of diabetic retinopathy. *Nature Medicine*, *30*(2), 584-594.
15. Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, *8*, 1-14.
16. Pan, L., Sun, W., Wan, W., Zeng, Q., & Xu, J. (2023). Research Progress of Diabetic Disease Prediction Model in Deep Learning. *Journal of Theory and Practice of Engineering Science*, *3*(12), 15-21.
17. Ahmad, H. F., Mukhtar, H., Alaqail, H., Seliaman, M., & Alhumam, A. (2021). Investigating health-related features and their impact on the prediction of diabetes using machine learning. *Applied Sciences*, *11*(3), 1173.
18. Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. *Journal of healthcare engineering*, *2021*(1), 9930985.
19. Nnamoko, N., & Korkontzelos, I. (2020). Efficient treatment of outliers and class imbalance for diabetes prediction. *Artificial intelligence in medicine*, *104*, 101815.
20. Kanbour, S., Harris, C., Lalani, B., Wolf, R. M., Fitipaldi, H., Gomez, M. F., & Mathioudakis, N. (2024). Machine learning models for prediction of diabetic microvascular complications. *Journal of diabetes science and technology*, *18*(2), 273-286.
21. Ismail, L., & Materwala, H. (2025). IDMPF: intelligent diabetes mellitus prediction framework using machine learning. *Applied Computing and Informatics*, *21*(1/2), 78-89.
22. Salih, M. S., Ibrahim, R. K., Zeebaree, S. R., Asaad, D., Zebari, L. M., & Abdulkareem, N. M. (2024). Diabetic prediction based on machine learning using PIMA indian dataset. *Communications on Applied Nonlinear Analysis*, *31*(5s), 138-156.
23. Natarajan, K., Baskaran, D., & Kamalanathan, S. (2025). An adaptive ensemble feature selection technique for model-agnostic diabetes prediction. *Scientific Reports*, *15*(1), 6907.
24. Hasan, M. K., Saeed, R. A., Alsuhibany, S. A., & Abdel-Khalek, S. (2022). An empirical model to predict the diabetic positive using stacked ensemble approach. *Frontiers in Public Health*, *9*, 792124.
25. Alnowaiser, K. (2024). Improving healthcare prediction of diabetic patients using KNN imputed features and tri-ensemble model. *IEEE Access*, *12*, 16783-16793.
26. Jose, R., Syed, F., Thomas, A., & Toma, M. (2024). Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions. *Applied Sciences*, *14*(5), 2132.
27. Teki, S., Sriharsha, K., & Nandimandalam, M. (2021). A diabetic prediction system based on mean shift clustering. *nutrition*, *84*, S177-S181.
28. Chatterjee, R., Akhtar, M. A. K., Pradhan, D. K., Chakraborty, F., Kumar, M., Verma, S., ... & García-Arenas, M. (2024). FNN for diabetic prediction using oppositional whale optimization algorithm. *IEEE Access*, *12*, 20396-20408.

29. Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). [Retracted] A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *Journal of healthcare engineering*, *2022*(1), 1684017.
30. Sheng, G., Kuang, M., Yang, R., Zhong, Y., Zhang, S., & Zou, Y. (2022). Evaluation of the value of conventional and unconventional lipid parameters for predicting the risk of diabetes in a non-diabetic population. *Journal of translational medicine*, *20*(1), 266.
31. Lama, L., Wilhelmsson, O., Norlander, E., Gustafsson, L., Lager, A., Tynelius, P., ... & Östenson, C. G. (2021). Machine learning for prediction of diabetes risk in middle-aged Swedish people. *Heliyon*, *7*(7).
32. Rajesh, G., Raajini, X. M., Sagayam, K. M., & Dang, H. (2020). A statistical approach for high order epistasis interaction detection for prediction of diabetic macular edema. *Informatics in Medicine Unlocked*, *20*, 100362.
33. Zhang, Y., Xu, F., Lin, Z., Wang, J., Huang, C., Wei, M., ... & Li, J. (2022). Prediction of Visual Acuity after anti-VEGF Therapy in Diabetic Macular Edema by Machine Learning. *Journal of Diabetes Research*, *2022*(1), 5779210.
34. Li, Z. Q., Fu, Z. X., Li, W. J., Fan, H., Li, S. N., Wang, X. M., & Zhou, P. (2023). Prediction of diabetic macular edema using knowledge graph. *Diagnostics*, *13*(11), 1858.
35. Fan, Y., Long, E., Cai, L., Cao, Q., Wu, X., & Tong, R. (2021). Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes. *Frontiers in pharmacology*, *12*, 665951.
36. Liu, B., Zhang, B., Hu, Y., Cao, D., Yang, D., Wu, Q., ... & Yu, H. (2021). Automatic prediction of treatment outcomes in patients with diabetic macular edema using ensemble machine learning. *Annals of Translational Medicine*, *9*(1), 43.
37. Sumathy, B., Chakrabarty, A., Gupta, S., Hishan, S. S., Raj, B., Gulati, K., & Dhiman, G. (2022). Prediction of diabetic retinopathy using health records with machine learning classifiers and data science. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, *11*(2), 1-16.
38. Cao, J., You, K., Jin, K., Lou, L., Wang, Y., Chen, M., ... & Ye, J. (2021). Prediction of response to anti-vascular endothelial growth factor treatment in diabetic macular oedema using an optical coherence tomography-based machine learning method. *Acta ophthalmologica*, *99*(1), e19-e27.
39. Mustafa, H. A., Albkrye, A. M. S., AbdAlla, B. M., Khair, M. A. M., Abdelwahid, N., & Elnasri, H. A. (2020). Computational determination of human PPARG gene: SNPs and prediction of their effect on protein functions of diabetic patients. *Clinical and Translational Medicine*, *9*, 1-10.
40. Soni, M., & Varma, S. (2020). Diabetes prediction using machine learning techniques. *International Journal of Engineering Research & Technology (IJERT)*, *9*(09), 2278-0181.
41. Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine learning models. *International Journal of Environmental Research and Public Health*, *19*(19), 12378.
42. 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.
43. Shin, J., Kim, J., Lee, C., Yoon, J. Y., Kim, S., Song, S., & Kim, H. S. (2022). Development of various diabetes prediction models using machine learning techniques. *Diabetes & Metabolism Journal*, *46*(4), 650-657.
44. Zhu, T., Li, K., Herrero, P., & Georgiou, P. (2022). Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Transactions on Biomedical Engineering*, *70*(1), 193-204.
45. Zhu, T., Kuang, L., Piao, C., Zeng, J., Li, K., & Georgiou, P. (2024). Population-specific glucose prediction in diabetes care with transformer-based deep learning on the edge. *IEEE transactions on biomedical circuits and systems*, *18*(2), 236-246.
46. Daliya, V. K., & Ramesh, T. K. (2025). A cloud based Optimized Ensemble model for risk prediction of diabetic progression-An Azure Machine Learning perspective. *IEEE Access*.
47. Shams, M. Y., Tarek, Z., & Elshewey, A. M. (2025). A novel RFE-GRU model for sdiabetes classification using PIMA Indian dataset. *Scientific Reports*, *15*(1), 982.
48. Khurshid, M. R., Manzoor, S., Sadiq, T., Hussain, L., Khan, M. S., & Dutta, A. K. (2025). Unveiling diabetes onset: Optimized XGBoost with Bayesian optimization for enhanced prediction. *PloS one*, *20*(1), e0310218.
49. Noh, M. J., & Kim, Y. S. (2025). Diabetes Prediction Through Linkage of Causal Discovery and Inference Model with Machine Learning Models. *Biomedicines*, *13*(1), 124.
50. Sajid, M., Malik, K. R., Khan, A. H., Iqbal, S., Alaulamie, A. A., & Ilyas, Q. M. (2025). Next-generation diabetes diagnosis and personalized diet-activity management: A hybrid ensemble paradigm. *PloS one*, *20*(1), e0307718.