

Early Prediction of Diabetes using Machine Learning and Deep Learning Techniques – A Comparative Analysis

J. Jannathul Firthous^{1,2}, Dr. G. Murugeswari³

¹Research Scholar, Reg No: 18221192162012, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India

²Assistant Professor, Department of Computer Science, Sadakathullah Appa College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli-627011, Tamil Nadu, India

³Associate Professor, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India

DOI: [https://doi.org/10.63001/tbs.2026.v21.i02.S.I\(2\).pp1076-1091](https://doi.org/10.63001/tbs.2026.v21.i02.S.I(2).pp1076-1091)

Keywords:

*Diabetes Prediction,
Machine Learning,
CNN,
LSTM, BiLSTM.*

Received on:

24-04-2026

Accepted on:

11-05-2026

Published on:

22-05-2026

Abstract

One of the most common chronic illnesses in the world, diabetes mellitus has grown in frequency and complications, making it a serious public health issue. Early detection and treatment of diabetes can enhance healthcare management and drastically lower death rates. Recently, methods for machine learning (ML) and deep learning (DL) have demonstrated encouraging results in disease prediction systems. Using the PIMA Indians Diabetes Dataset, this work compares a few of machine learning and deep learning methods for diabetes prediction. The suggested comparative framework assesses machine learning techniques like Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) and deep learning techniques like Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM). Results from experiments show that deep learning models perform better in prediction than traditional ML approaches. Our analysis found that the BiLSTM model performed about 98.7% better than the other algorithms.

1. INTRODUCTION

Diabetes mellitus is a long-term metabolic condition caused by elevated blood sugar levels due to inadequate insulin synthesis or the body's improper use of insulin. Global health reports state that during the past few decades, the prevalence of diabetes has significantly increased, placing a notable strain on healthcare systems across the globe. To reduce serious complications and increase patient survival

rates, diabetes must be detected early. Early detection may be delayed by traditional diagnostic methods, which frequently call for laboratory testing and ongoing observation.

Machine learning and deep learning approaches have become viable options for automated diabetes prediction with the development of artificial intelligence (AI). These methods can find hidden patterns

linked to diabetic problems by analyzing patient health indicators, physiological signals, and medical records. The ease of use, interpretability, and effectiveness of machine learning algorithms in managing structured datasets make them popular. Conversely, deep learning techniques may automatically extract intricate features from medical data without the need for human feature engineering. However, the performance of these techniques varies depending on data size, complexity, feature distribution, and preprocessing methods. Therefore, a comparative study is necessary to evaluate the strengths and limitations of different algorithms. Several researchers have explored ML and DL approaches for diabetes prediction using datasets such as the Pima Indian Diabetes Dataset (PIDD), Electronic Health Records (EHR), and physiological signal datasets. Using the Pima dataset, this study compares the performance of Machine Learning techniques including Logistic Regression, Support Vector Machine, and XGBoost with Deep Learning models like CNN, LSTM, and BiLSTM.

This paper is organized as follows: Section 2 presents the related work on diabetes prediction using machine learning and deep learning approaches. Section 3

describes the machine learning methods including Logistic Regression, Support Vector Machine and XGBoost also the deep learning methods such as CNN, LSTM and BiLSTM used in the study. Section 4 discusses the experimental results and comparative analysis of the applied models. Finally, Section 5 concludes the study and highlights future research directions.

2. RELATED WORK

Conventional diabetes detection methods such as Fasting Blood Glucose (FBG), Oral Glucose Tolerance Test (OGTT), HbA1c, and Random Glucose Testing provide structured clinical information for diagnosis. However, machine learning and deep learning techniques improve early diabetes prediction by identifying hidden patterns in patient data and supporting clinical decision-making.

Ahmed T. [1] applied machine learning models for identifying high-risk diabetic patients and it was reported that Support Vector Machine (SVM) achieved 77% - 79% of performance after preprocessing and evaluated using accuracy, precision, recall, and F1-score. Similarly, Wankhede, V.A et al. [2] compared various machine learning algorithms including

SVM, KNN, ANN, Naive Bayes, Logistic Regression, and Decision Trees on the Pima Indian Diabetes Dataset, concluding that SVM produced the accuracy of 81.3%. Clara Ribeiro Neves et al. [3] utilized physiological signals such as HR, RR, and SpO₂ with wavelet decomposition and Linear SVM, and achieved 94% of accuracy and improved recall for early diabetes detection. Narasimharao M. et al. [4] emphasized hybrid machine learning approaches using algorithms such as SVM and Naive Bayes, and reported the accuracy levels ranging from 65% to 97%. Tared, S et al. [5] developed a hybrid Artificial Bee Colony optimized SVM model, by combining optimization and classification techniques to enhance the accuracy of 92.47%. Jena, S.K et al. [6] developed a diabetes prediction framework using preprocessing methods such as KNN Imputer, SMOTE, and Recursive Feature Elimination, where an ensemble voting classifier achieved the best accuracy of 84.2%. Jain U. et al. [7] highlighted the effectiveness of SVM for accurate diabetes classification and healthcare decision support. Alzboon M.S et al. [8] evaluated several algorithms including Logistic Regression, Random Forest, SVM, Gradient Boosting, and Neural Networks on the Pima

dataset, where Neural Networks achieved the highest accuracy of 78.57%. Logistic Regression models [9][10] were widely used as interpretable baseline classifiers, identifying glucose, BMI, and hereditary factors as significant predictors with an accuracy of 78.3%.

Zaki, M.G et al. [11] used the CDC BRFSS 2015 dataset with SMOTE and feature selection techniques, where Logistic Regression achieved a recall of 75.06% and ROC-AUC of 83.62% compared to Random Forest and XGBoost. Abu-Shareha et al. [12] developed a hybrid clustering and feature selection framework, where Random Forest and XGBoost achieved 88.5% accuracy. Martinović G. et al. [13] demonstrated the importance of preprocessing techniques such as median imputation, and outlier handling for improving diabetes prediction performance. XGBoost-based approaches with explainable AI methods such as SHAP and LIME were introduced by Rafie, Z et al. [14] for type-2 diabetes prediction. Khurshid, M.R et al. [15] enhanced XGBoost performance using Bayesian optimization, achieving 94.26% accuracy. Mazgouti, L et al. [16] combined XGBoost and LSTM for blood glucose prediction using Continuous Glucose Monitoring

(CGM) data. Jawza D.N. et al. [17] applied feature selection methods such as Genetic Algorithm and Particle Swarm Optimization to improve the performance of Random Forest and XGBoost performance. Swapna Donepudi et al. [18] compared LightGBM, XGBoost, and hybrid models, where LightGBM achieved the best accuracy of 95.43%.

Recent studies have increasingly focused on deep learning techniques for diabetes prediction. Sing, K.R et al. [19] designed a CNN-LSTM framework with Bayesian optimization and Conditional GAN-based augmentation, achieving 94% accuracy and 98% AUC. Yahyaoui, A, et al. [20] Predicted the diabetes using Machine Learning and Deep Learning algorithms based on patient medical data. The algorithms used are Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN), which help healthcare professionals improve early diagnosis and clinical decision making. Among the tested algorithms, Random Forest achieved the highest accuracy of 83.67%, followed by CNN with 76.81% and SVM with 65.38%, showing that Random Forest performed best for diabetes prediction.

Mulla, R.M et al. [21] implemented a hybrid CNN-LSTM architecture with Random Forest feature selection and SMOTE, outperforming conventional approaches with recall and F1-score of 89%. Zhou, H. et al. [22] achieved 99.41% accuracy on the Pima dataset using deep learning classification techniques. Alex, S.A et al. [23] used Deep 1D CNN with SMOTE and outlier detection for improved diabetes classification on imbalanced datasets. Hybrid recurrent neural network models have also shown promising results. Raja, W et al. [24] developed LSTM-RNN with Grey Wolf Optimization based feature selection for diabetes prediction achieved with 95.21% accuracy. Imam A.A. et al. [25] designed a BiLSTM-BiGRU hybrid model that captured complex medical data patterns for type-2 diabetes prediction and achieved 98% of accuracy. Temporal Attention-Gated BiLSTM models [26] utilized longitudinal Electronic Health Records (EHR) for sequential diabetes prediction and attained ~70% of AUROC compared to standard LSTM models. Singh et al. [27] developed DiaXplain, a CNN-XGBoost hybrid framework for Type-2 diabetes diagnosis that achieved 98.24% accuracy, 95.12% precision, and a 97.50% F1-score on NHANES and PIMA

datasets. The study integrates SHAP-based explainability to provide transparent, patient-level and global interpretations of diagnostic predictions, improving the clinical reliability of AI-assisted diabetes detection. Prabhu, P et al. [28] developed a deep belief neural network (DBNN) model for diabetes mellitus prediction that achieved 96.3% accuracy on the PIMA Indian Diabetes dataset. The multilayer deep learning architecture improved diabetes classification performance through hierarchical feature extraction and optimized prediction capability.

Advanced hybrid architectures further improved the performance and interpretability. Suyanto et al. [29] introduced a Multi-Voter Multi-Commission Nearest Neighbor deep learning framework for diabetes diagnosis, while Chun Ouyang et al. [30] employed the CNN models on peripheral pulse wave

data for low-cost diabetes screening. Osama Rabie et al. [31] developed a BiLSTM-based Decision Support System for diabetes diagnosis using patient information. Marmik Shrestha et al. [32] integrated SVM, Radial Basis Function, and LSTM for predicting Type-2 diabetes prediction. Rajan R. et al. [33] developed the BiTabNet_CFs framework incorporating BiLSTM, TabNet, CTGAN, and counterfactual explanations, achieving 98.7% accuracy with enhanced interpretability. Liu, L [28] developed an MHA-CNN-BiLSTM-based diabetes prediction model using digestive health factors that achieved 98.7% accuracy with improved feature representation and temporal learning capability. The study incorporated interpretability analysis to identify the contribution of digestive-related attributes in diabetes prediction and enhance model transparency.

3. METHODOLOGY

In this work, a comparative analysis of few machine Learning techniques and deep

learning techniques were carried out. The flow diagram is presented in figure 3.1

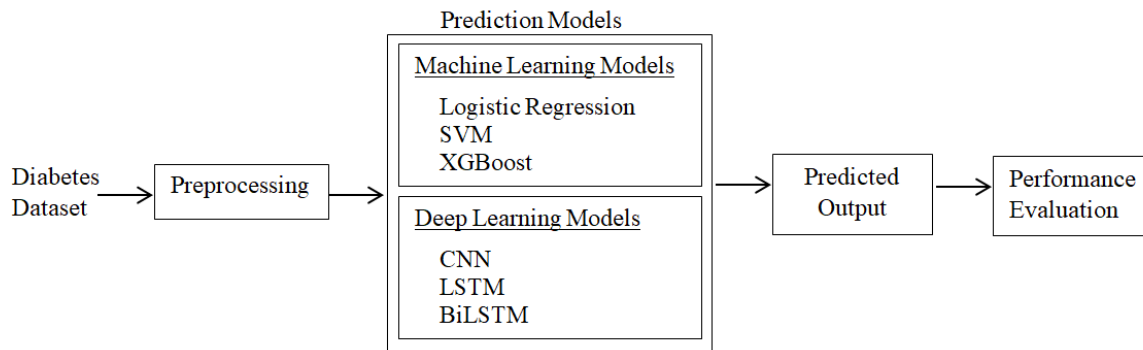


Figure 1 Flow diagram of Machine Learning and Deep Learning techniques for comparison

3.1 Preprocessing

Pre-processing techniques play a crucial role in Diabetes Mellitus prediction by enhancing the quality and consistency of medical datasets before applying machine learning and deep learning algorithms. The techniques include data cleaning to eliminate noise and inconsistencies, missing value imputation using mean or median replacement, and normalization to scale data into a uniform range.

3.2 Machine Learning Models

Machine Learning algorithms are capable of identifying patterns from historical medical data and generating predictions based on learned relationships. These methods require structured datasets with predefined features such as glucose level, blood pressure, insulin level, BMI, age, and skin thickness.

3.2.1 Logistic Regression

Logistic Regression is one of the most commonly used supervised machine learning algorithms for classification problems, especially binary classification tasks such as diabetes prediction. The algorithm estimates probability of diabetes occurrence using a sigmoid activation function. The model first computes a weighted sum of input features given in equation (1). Logistic Regression first calculates a weighted sum of the input features.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_n x_n \quad (1)$$

Where z represents the weighted sum of input, β_0 is the bias term, $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients and $x_1, x_2, \dots, x_n$ are the features. The calculated value is then passed through the sigmoid function to convert it into probability as expressed in equation (2).

$$P = \frac{1}{1+e^{-z}}$$

(2)

Here, P represents the probability that a patient has diabetes, while z represents the weighted combination of input features. If the predicted probability falls between 0 and 1, with values nearer 1 denoting the patient is classified as diabetic.

3.2.2 Support Vector Machine (SVM)

Support Vector Machine is a supervised machine learning algorithm widely used for classification tasks. In diabetes prediction, SVM is primarily applied as a binary classification model to distinguish diabetic patients from non-diabetic individuals based on clinical features. The main objective of SVM is to identify an optimal hyperplane that separates the data points of different classes with the maximum possible margin. A larger margin between classes improves the model's generalization ability and reduces classification errors. The hyperplane used in SVM is represented in equation (3)

$$w^T \cdot x + b = 0$$

(3)

where w represents the weight vector, x represents the input feature vector, and b

denotes the bias term. The classification function is expressed in equation (4)

$$f(x) =$$

$$\text{sign}(w^T \cdot x + b)$$

$$(4) \quad \text{where} \quad f(x)$$

determines the class label of the input data.

If the output is positive, the sample is classified as diabetic otherwise it is classified as non-diabetic.

3.2.3 Extreme Gradient Boosting (XGBOOST)

Extreme Gradient Boosting, commonly known as XGBoost, is an advanced ensemble machine learning algorithm based on gradient boosting decision trees. In diabetes prediction, XGBoost analyzes patient medical attributes from the Pima Indian Diabetes Dataset to classify patients as diabetic or non-diabetic. The algorithm combines multiple weak decision tree models sequentially to create a strong predictive model capable of minimizing classification errors. The objective function of XGBoost is represented in equation (5)

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i)$$

$$+ \sum_{k=1}^K \Omega(f_k) \quad (5)$$

where $l(y_i, \hat{y}_i)$ denotes the loss function that measures the difference between the

actual and predicted values. Decision trees are constructed iteratively by XGBoost, with each new tree aiming to fix the mistakes made by earlier trees. This mechanism improves classification accuracy and enhances robustness against overfitting.

3.3 Deep Learning Models

Deep learning techniques have gained significant attention in diabetes prediction due to their ability to identify complex patterns and relationships within medical data. These methods are applied using the PIMA Indian Diabetes Dataset to develop predictive models that support accurate and early detection of diabetes.

3.3.1 Convolutional Neural Network (CNN):

In diabetes prediction, the Convolutional Neural Network (CNN) architecture consists mainly of Convolutional layers, Activation layers, Pooling layers, Flatten layers, and Fully Connected layers. The Convolutional layer extracts important diabetic feature patterns from clinical attributes using learnable kernels. These kernels generate feature maps that capture hidden diabetic correlations from the input data. After convolution, the ReLU activation function

is applied to remove negative activations and improve nonlinear learning capability, allowing the model to identify complex diabetic relationships more effectively. The Pooling layer mainly uses Max Pooling with a 2×2 pooling window to reduce feature map dimensions while preserving the strongest diabetic feature responses. This process decreases computational complexity and minimizes overfitting during training. The Flatten layer converts pooled multidimensional feature maps into one-dimensional vectors, which are forwarded into the Fully Connected layer for classification. Finally, the Fully Connected layer uses Sigmoid activation to compute probability scores and classify patients as diabetic or non-diabetic based on the extracted clinical feature representations.

3.3.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized Recurrent Neural Network architecture used for learning long-term dependencies and sequential patterns from diabetes-related medical data. LSTM processes patient records sequentially and captures relationships among clinical parameters to identify diabetic conditions. Unlike traditional recurrent networks,

LSTM effectively handles long-range dependencies without suffering from vanishing gradient problems.

The LSTM architecture mainly consists of Memory Cells, Forget gates, Input gates, Output gates, and Fully Connected layers. The Forget gate removes unnecessary clinical information from memory, the Input gate updates important diabetic feature information into the cell state, and the Output gate generates hidden feature representations for prediction. The learned diabetic feature patterns are forwarded into Fully Connected layers for classification. Finally, the Sigmoid activation function is applied in the output layer to classify patients as diabetic or non-diabetic based on the extracted sequential clinical patterns.

3.3.3 Bidirectional Long Short-Term Memory (BiLSTM)

Bidirectional Recurrent Neural Networks, commonly known as BiLSTM, is an enhanced version of LSTM that processes sequential information in both forward and backward directions. Unlike standard LSTM, which learns information only from past sequences, BiLSTM captures dependencies from both previous and future time steps simultaneously. This

bidirectional learning capability enables the model to achieve better contextual understanding and improved feature representation.

The BiLSTM architecture mainly consists of Forward LSTM layers, Backward LSTM layers, Memory Cell states, Forget gates, Input gates, Output gates, and Fully Connected layers. The Forget gate removes irrelevant clinical information, the Input gate updates important diabetic feature information into memory cells, and the Output gate generates relevant hidden representations for prediction. After bidirectional sequence learning, the extracted feature representations are passed into Fully Connected layers for classification. Finally, the Sigmoid activation function is used in the output layer to classify patients as diabetic or non-diabetic based on learned sequential diabetic patterns.

4. EXPERIMENTAL ANALYSIS

In this work, the experimental analysis evaluates the effectiveness of the machine learning and deep learning models using the PIMA datasets through standard performance and the comparative analysis.

4.1 Dataset

The Pima Indian Diabetes consists

of 768 patient records with medical attributes such as glucose level, BMI, insulin concentration, blood pressure, pregnancies, skin thickness, diabetes pedigree function, and age. Before training the models, preprocessing techniques including missing value handling, feature normalization, and dataset balancing were applied to improve prediction performance and reduce data inconsistencies. The dataset was divided into training and testing subsets using 80:20 ratio.

4.2 Performance Metrics

The metrics like Accuracy, Recall, Precision, and F1-Score to evaluate the models performance.

Accuracy:

Accuracy is the proportion of correctly classified instances out of total number of instances, show-casing overall performance but potentially missing class balance in imbalanced datasets. The equation is as follows

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

where TP: True Positives; TN: True Negatives; FP: False Positives; FN: False Negatives.

Precision:

Precision is defined as the percentage of accurately anticipated positive instance among all positive instances. It is illustrated in the given equation

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall:

Recall, also known as sensitivity or true positive rate, measures the models capacity in detecting all positive instances. It is crucial for tasks where missing positive cases has severe consequences, such as medical diagnoses.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score:

The F1-Score is the harmonic mean of Precision and Recall, providing a single metric to balance the trade-off between these two metrics. It is particularly useful when dealing with imbalanced datasets.

$$F1 - Score = 2 * \frac{Precision.Recall}{Precision+Recall} \quad (9)$$

4.3 Performance Analysis

The comparative performance of machine learning and deep learning models shown in table 1 was evaluated using standard

evaluation metrics including Accuracy, Precision, Recall, F1-Score.

Table 1 Comparative Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	78.4	67.0	64.0	65.5
Support Vector Machine	83.1	78.4	72.2	75.2
XGBoost	91.3	90.6	89.3	89.9
CNN	84.5	82.6	80.2	81.4
LSTM	93.4	92.8	92.5	92.6
BiLSTM	98.7	97.0	98.8	97.9

4.4 Discussion

The comparative performance analysis of Logistic Regression, Support Vector Machine, Extreme Gradient Boosting, Convolutional Neural Network, Long Short-Term Memory, and Bidirectional Long Short-Term Memory using the Pima Indian Diabetes Dataset revealed that Deep Learning approaches achieved better prediction performance compared to traditional Machine Learning models. Logistic Regression produced the lowest accuracy of 78.4% with moderate precision, recall, and F1-score values because of its limited capability to model nonlinear relationships among medical features. SVM improved the classification performance to 83.1% by effectively handling nonlinear

feature distributions using kernel functions. CNN achieved 84.5% accuracy through automatic feature extraction, while LSTM further improved performance to 93.4% by learning temporal dependencies and hidden feature patterns within healthcare data. Among all evaluated models, XGBoost and BiLSTM achieved the best performance with accuracies of 91.3% and 98.7%, respectively. XGBoost demonstrated strong classification capability because of its gradient boosting mechanism and effective feature interaction learning, whereas BiLSTM achieved the highest recall and AUC performance through bidirectional sequence learning and enhanced contextual understanding. The results indicate that

advanced Deep Learning and boosting algorithms provide superior diabetes prediction accuracy, although they require higher computational resources and training complexity compared to traditional Machine Learning methods.

5. CONCLUSION

In this study, a comparative analysis of Machine Learning Models such as Logistic Regression, Support Vector Machine, Extreme Gradient Boosting and Deep Learning Models such as Convolutional Neural Network, Long Short-Term Memory, and Bidirectional Recurrent Neural Networks was performed using the Pima Indian Diabetes Dataset for diabetes prediction. Among all evaluated classifiers, BiLSTM achieved the best overall performance with the highest accuracy, precision, recall, and F1-score, followed by XGBoost. The study concludes that Deep Learning approaches are more effective for intelligent diabetes prediction systems, although they require higher computational resources and longer training time. Overall, the findings highlight the importance of selecting appropriate predictive models based on dataset characteristics, computational requirements, and healthcare application needs for improving early

diabetes diagnosis and clinical decision support systems. In order to further enhance prediction accuracy and dependability in medical diagnosis systems, future research may concentrate on integrating deep learning architectures, larger real-world healthcare datasets, explainable artificial intelligence methodologies, and optimization algorithms.

REFERENCES

1. Ahmed, T. (2026). Diabetes Prediction using SVM.
2. Wankhede¹, V. A., More, A. R., Desai, P. S., Thakre, N. R., & Bachhav, S. M. (2026). Diabetes Prediction Through Machine Learning and Ontology.
3. Natekar, S., Singh, A., Pillai, N., & Hannah Grace, G. (2026). Diabetic Retinopathy Detection Using SVM. *Digital Cities*, 129-144.
4. Narasimharao, M., Swain, B., Priyadarshi, R., Nayak, P. P., & Bhuyan, S. (2025). A survey of machine learning techniques for diabetes prediction: current trends and future directions. *Archives of Computational Methods in Engineering*, 1-36.
5. Tared, S., Roubehie Fissa, M., Khaouane, L., & Hanini, S. (2025). Prediction of diabetes using hybrid

- support vector machines with artificial bee colony. *International Journal of Diabetes in Developing Countries*, 1-16.
6. Jena, S. K., Behera, D. K., Jena, A. K., & Rout, J. K. (2025). A Novel Ensemble Machine Learning Framework for Improved Diabetes Prediction and Complication Prevention. *Procedia Computer Science*, 258, 4008-4017.
7. Jain, U., Jain, D., & Rana, Y. V. (2025, June). Enhancing Healthcare Diagnostics with SVM-Based Machine Learning for Diabetes Prediction. In *International Conference on Advances and Applications in Artificial Intelligence (ICAAAI 2025)* (pp. 914-924). Atlantis Press.
8. Alzboon, M. S., Alqaraleh, M., & Al-Batah, M. S. (2025). Diabetes prediction and management using machine learning approaches. *arXiv preprint arXiv:2506.11501*.
9. Jithendra, V. & Mohit, R & Madhusudhan, M & Jagadeesh, B & Kusuma, S. (2023). Diabetes Prediction using Machine Learning Techniques. *Journal of Artificial Intelligence and Capsule Networks*. 5. 190-206. 10.36548/jaicn.2023.2.008.
10. Okegbade, A. I., & Egbewumi, A. O. (2026). Predicting Diabetes Mellitus Using Logistic Regression On Clinical And Demographic Data. *International Journal of Interdisciplinary Research*, 2(2), 152-173.
11. Zaki, M. G., & Imaduddin, H. (2026). Early Classification of Diabetes Risk in Productive Age Groups Using Machine Learning. *Jurnal Penelitian Sains Teknologi*, 75-90.
12. Abu-Shareha, A. A., Abualhaj, M., Hussein, A. H., Amer, A., Achuthan, A., & Halin, A. A. (2026). Diabetes Prediction Using Hybrid Supervised and Unsupervised Techniques Based on PIMA Dataset. *Journal of Artificial Intelligence and Technology*, 6, 79-87.
13. Martinović, G., Ivković, I., Verber, D., & Bačun, T. (2026). Effect of Data Preparation on Machine Learning Models for Diabetes Prediction. *Engineering Proceedings*, 125(1), 13.
14. Rafie, Z., Talab, M. S., Koor, B. E. Z., Garavand, A., Salehnasab, C., & Ghaderzadeh, M. (2025). Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes. *BMC Public Health*, 25(1), 3688.

15. 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.
16. Mazgouti, L., Laamiri, N., Ali, J. B., El Idrissi, N. E. A., Di Costanzo, V., Naeck, R., & Ginoux, J. M. (2025). Optimization of blood glucose prediction with LSTM-XGBoost fusion and integration of statistical features for enhanced accuracy. *Biomedical Signal Processing and Control*, 107, 107814.
17. Jawza, D. N., Mazdadi, M. I., Farmadi, A., Saragih, T. H., Kartini, D., & Abdullayev, V. (2025). The enhancing diabetes prediction accuracy using random forest and XGBoost with PSO and GA-based feature selection. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(2), 295-306.
18. Donepudi, S., Nakka, R., Thota, K. K., Ajmeera, M., Praveen, S. P., & Sindhura, S. (2025). Enhancing Person-Centric Health Care for Diabetes Prediction: A Comparative Study of LightGBM, XGBoost, and Hybrid LIGB Model. In *Enabling Person-Centric Healthcare Using Ambient Assistive Technology, Volume 2: Personalized and Patient-Centric Healthcare Services in AAT* (pp. 127-155). Cham: Springer Nature Switzerland.
19. Singh, K. R., Dash, S., Liu, H., & Wang, Z. (2026). Enhanced diabetes prediction using pre-trained CNNs, LSTM, and conditional GAN on transformed numerical data. *Scientific Reports*.
20. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. In *2019 1st International informatics and software engineering conference (UBMYK)* (pp. 1-4). IEEE.
21. Mulla, R. M., & Desai, K. R. (2026). A Hybrid CNN-LSTM Framework for Diabetes Risk Prediction in COVID-19 Patients. *Journal homepage: <http://iieta.org/journals/isi>*, 31(2), 499-510.
22. Zhou H, Myrzashova R, Zheng R (2020) Diabetes prediction model based on an enhanced deep neural network. *J Wirel Com Netw* 148(2020).
23. Alex, S. A., Nayahi, J. J. V., Shine, H., & Gopirekha, V. (2022). Deep convolutional neural network for

- diabetes mellitus prediction. *Neural Computing and Applications*, 34(2), 1319-1327.
24. Raja, W., Chandrababha, K., Ruckmani3b, V., & Gowri4c, S. (2026). Implementing LSTM-RNN for improved diabetic dataset classification.
 25. Imam, A. A., Junaidu, S. B., Mamman, H., Kumar, G., Balogun, A. O., Khowaja, S. A., ... & Conteh, F. (2026). DeepClassifier: A Data Sampling-Based Hybrid BiLSTM-BiGRU Neural Network for Enhanced Type 2 Diabetes Prediction. *Computer Modeling in Engineering & Sciences*, 146(3).
 26. Rao, M. M., Manjunath, C., Venkatesh, D., & Vasundhara, S. (2026, February). Longitudinal Risk Prediction of Hospital Readmission in Diabetes Management Using a Temporal Attention-Gated BiLSTM with Focal Loss. In 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 230-233). IEEE.
 27. Singh S, Wani NA, Kumar R, Bedi J. DiaXplain: a transparent and interpretable artificial intelligence approach for Type-2 diabetes diagnosis through deep learning. *Comput Electr Eng*. 2025;126:110470. doi:10.1016/j.compeleceng.2025.110470.
 28. Prabhu P, Selvabharathi S. Deep belief neural network model for prediction of diabetes mellitus. In: *Proceedings of the 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*. New York, NY, USA: ACM; 2019. p. 138–42.
 29. Suyanto, S., Meliana, S., Wahyuningrum, T., Khomsah, S. (2022). A new nearest neighbor-based framework for diabetes detection. *Expert Systems with Applications*, 199, 116857. <https://doi.org/10.1016/j.eswa.2022.116857>.
 30. Ouyang, C., Gan, Z., Zhen, J., Guan, Y., Zhu, X., Zhou, P. (2021). Inter-patient classification with encoded peripheral pulse series and multi-task fusion cnn: application in type 2 diabetes. *IEEE journal of biomedical and health informatics*, 25(8), 3130-3140. <https://doi.org/10.1109/JBHI.2021.3061114>.
 31. Rabie, O., Alghazzawi, D., Asghar, J., Saddozai, F.K., Asghar, M.Z. (2022). A decision support system for diagnosing diabetes using deep neural network. *Frontiers in public health*, 10, 861062.

32. Shrestha, M., Alsadoon, O.H., Alsadoon, A., Al-Dala'in, T., Rashid, T.A., Prasad, P.W.C., Alrubaie, A. (2023). A novel solution of deep learning for enhanced support vector machine for predicting the onset of type 2 diabetes. *Multimedia Tools and Applications*, 82(4), 6221-6241.
33. Rajan, R., C. Karthik, J. Gnana Jeslin, and Nagarajan Gurusamy. "BiTabNet_CFs: A Novel Hybrid Deep Learning Framework for Diabetes Mellitus Detection." *International Journal of Pattern Recognition and Artificial Intelligence* (2026): 2540007.
34. Liu, L. (2026, March). Research on Diabetes Prediction Model Based on Digestive Factors Using MHA-CNN-BiLSTM and Interpretability Analysis. In 2026 9th International Symposium on Big Data and Applied Statistics (ISBDAS) (pp. 191-194). IEEE.