Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 2, May 2026, pp. 509~517
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i2.pp509-517

An enhanced deep learning model with context-aware attention for diabetes prediction

J. Jannathul Firthous[1,2], G. Murugeswari[1]
[1]Department of Computer Science and Engineering, School of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India
[2]Department of Computer Science, Sadakathullah Appa College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, India

Article Info

Article history:

Received Mar 29, 2025
Revised Feb 24, 2026
Accepted May 26, 2026

Keywords:

BiLSTM
CNN
Deep learning
Diabetes prediction
Multi-layer attention mechanism

ABSTRACT

A plethora of people worldwide suffer from diabetes, a chronic, potentially fatal illness that resulted serious risks and complications if left untreated. Effective management requires early prediction and intervention. Despite their advantages, traditional machine learning techniques frequently find it difficultly in grasping the intricate temporal as well as geographical correlations included with in medical stats. For the purpose of effectively forecast diabetes mellitus, the proposed work suggests a unique deep learning model called multilayer diabetes deep learning attention with context mechanism (MLDDAM). This model incorporates a hybrid architecture that integrates an Attention with Context Mechanism to enhance the model's efficiency will be conversant with emphasizing on key aspects, convolutional neural networks (CNN) are utilized to extract traits, and bidirectional long short-term memory (BiLSTM) captures sequential dependencies. This innovative design enables the model to perform better by utilizing the input data's temporal and geographical properties. Experiments using benchmark datasets show that the suggested MLDDAM model is efficient and robust, with outstanding 99.43% prediction accuracy for diabetes. These outcomes demonstrate the MLDDAM model's effectiveness as a precise and dependable tool to assist clinical decision-making in the management of diabetes.

Corresponding Author:

J. Jannathul Firthous
Department of Computer Science and Engineering, School of Computer Science and Engineering
Manonmaniam Sundaranar University
Abishekapatti, Tirunelveli-627012, Tamil Nadu, India
Email: janfirth1@gmail.com

1. INTRODUCTION

Glucose Disregulation Disorder, otherwise known as diabetes mellitus, ranks among the most prevalent metabolic conditions, which is a significant reason for morbidity and death, affecting millions of people globally. It is marked by high blood sugar as the human's the body cannot effectively manage insulin. The two most common types of diabetes are Type 1 and Type 2; Type 2 diabetes is more common and is mostly linked to lifestyle factors like poor diet and inactivity [1]. The World Health Organization (WHO) estimates that by 2030, around 500 million people worldwide will have diabetes, a condition that has been rapidly rising in prevalence [2], [3].

Blood glucose readings, such as those from the HbA1c test, postprandial glucose levels, and fasting blood sugar (FBS), are commonly used in traditional diabetes diagnosis techniques. Nevertheless, these tests are not sufficient to identify diabetes in its early stages, especially in those who may not exhibit any symptoms but are at risk. Furthermore, although diabetes prediction has been done using traditional machine learning techniques like logistic regression, decision trees, and support vector machines. The specified algorithms frequently fall short due to the temporal and complicated nature of the health data. Because of this, precisely forecasting diabetes, particularly in its early stages, continues to be a challenge. The advent of deep learning has created new opportunities for addressing these critical issues. Deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated significant promise in the analysis of huge, complicated datasets. While RNNs, especially long short-term memory (LSTM) networks, could be well appropriate to process sequential data, time-series data from continuous glucose monitoring devices, CNNs are good at extracting spatial characteristics from structured data, like medical records. Additionally, attention mechanisms [4] have been incorporated into deep learning models to enhance their interpretability and concentrate on the most essential aspects for making decisions [5].

Although these developments show promise, they frequently rely on static information or overlook the dynamic, real-time nature of managing diabetes. As wearable technology and continuous glucose monitoring (CGM) systems become more widespread. The requirement for data processing models is on the rise in real time as well as it makes predictions continuously. This requires models that can manage massive data sets and adjust to patients' evolving health over time.

To tackle these issues, the proposed study focused on a peculiar deep learning architecture, the multi-layer diabetes deep learning attention with context mechanism (MLDDAM), which incorporates the power of CNN, bidirectional long short-term memory (BiLSTM), and an attention mechanism to predict diabetes with high accuracy. By assimilating both spatial and temporal dependencies in the data, the objective of the MLDDAM model is to bring forth a more authentic, dynamic, and interpretable prediction system for early diabetes detection.

The research work is motivated by the increasing predominance of diabetes globally, a condition that has serious health concerns, such as renal failure, nerve damage, and heart disease. Early diagnosis is still a dominant obstacle despite improvements in diabetes care because conventional diagnostic techniques pre-dominantly abandon to capture the disease's complexity. The materialistic and dynamic nature of diabetes is strenuous for current machine learning models to handle, especially in real-time monitoring settings where wearable technology and continuous glucose monitoring (CGM) systems are becoming more and more common. Furthermore, there is a surging urgency for predictive models that help medical practitioners make well-informed judgments by providing both high accuracy and interpretability. This research seeks to fill these voids by proposing an exclusive deep learning model, MLDDAM, which combines CNN, BiLSTM, and attention mechanisms to provide highly accurate, real-time diabetes prediction with enhanced interpretability, ultimately improving early detection, patient outcomes, and proactive management of the disease.

The proposed model is designed to address assorted key challenges:

- Spatial and temporal feature extraction: by integrating CNN and BiLSTM, the model can extract both spatial and temporal patterns in patient health data, such as glucose levels and other relevant biomarkers, which are crucial for accurate diabetes prediction.
- Attention mechanism for interpretability: the attention with context mechanism dynamically weighs the implication of different features, improving the interpretability of the model's predictions and making it easier for healthcare providers to understand the reasoning behind the model's output.
- Improved prediction accuracy: the MLDDAM model demonstrates outstanding performance, achieving a 99.43% prediction accuracy for diabetes classification. This high accuracy shows the model's potential to provide reliable and precise predictions, which can significantly assist in early detection and intervention for diabetes.
- Validation on benchmark datasets: extensive experiments conducted on benchmark datasets confirm the robustness and efficiency of the proposed model, showcasing its capability as a reliable solution for real-world applications.

The study's framework is organized as follows: Section 2 reviews the literature survey, whereas Section 3 outlines the methodology, including a description of the Pima Indians Diabetes Database (PIMA-IDD), the models employed for diabetes diagnosis, and the proposed framework for diabetes prediction. Section 4 presents the results and discussions, and Section 5 concludes the paper by providing a future direction for research.

2. RELATED WORK

Many approaches have been suggested and documented for diagnosing diabetes. One notable machine learning framework, presented in [6], implemented various classification algorithms, including linear discriminant analysis (LDA) [7], quadratic discriminant analysis (QDA) [8], naive Bayes (NB) [9], Gaussian process classification (GPC) [10], support vector machine (SVM) [11], artificial neural network (ANN) [12], AdaBoost (AB) [13], logistic regression (LR) [14], decision tree (DT) [15], and random forest (RF) [16]. This framework incorporated various dimensionality reduction techniques and cross-validation methods to enhance model performance. Additionally, it involved experiments on outlier rejection and handling missing data, achieving a maximum area under the curve (AUC) of 0.930. Sonar and Jayamalini [17] implied a diabetes prediction model using machine learning algorithms such as decision tree, ANN, naive Bayes, and SVM. Among these, the decision tree algorithm achieved a higher precision of 85%. Similarly, Wei *et al.* [18] steered a predictive model utilizing algorithms like naive Bayes, deep neural networks (DNN), logistic regression, and decision trees. The above-mentioned study divulged that the model attained enhanced fidelity DNN model achieved the highest accuracy of 77.86%, surpassing the other algorithms included in the research.

Deep learning techniques have gained considerable traction recently in diagnosing and predicting diabetes mellitus. Convolutional neural networks (CNNs) were used by García-Ordás *et al.* [19] to predict diabetes; the PIMA-IDD dataset was used to equip their model. Similar to this, Khanam and Foo [20] investigated DM prediction using machine learning and artificial neural network (ANN) techniques. According to their research, support vector machines (SVM) and reinforcement learning (RL) worked well hand in hand to predict DM. Subsequently, their two-layered ANN model has an accuracy of 88.6%. Wearable technology and systems are growing in popularity. Furthermore, there is a mounting need for predictive models that help medical practitioners make well-informed judgments by providing both high accuracy and interpretability.

In order to categorize data from the PIMA-IDD dataset, Kannadasan *et al.* [21] utilized a deep ANN with a stacked encoder, demonstrating its promise for diabetes prediction. Swapna *et al.* [22] employed deep learning techniques to distinguish between normal and diabetic heart rate variability (HRV) signals. They created models based on CNNs and LSTM networks. They pulled out intricate temporal dynamics from HRV records, proving that deep learning approaches can handle diabetes-related time-series data. Naveena and Bharathi [23] amalgamate both moth flame optimization (MFO) and Crow search algorithm (CSA) to enhance deep learning architectures. This hybrid optimization technique has indicated its efficacy in refining the hyperparameters and training processes of deep learning models.

A CNN model was investigated by Zhu *et al.* [24] with the goal of forecasting future blood glucose levels in individuals with type 1 diabetes. They employed a refined edition of WaveNet, a popular model for processing acoustic signals. In order to determine the best deep learning algorithm, Kowsher *et al.* [25] have generated a deep neural network in conjunction with machine learning classifiers, assessing performance parameters like accuracy and precision. Similarly to this, Soniya *et al.* [26] recommended a hybrid evolutionary method that was merged with a CNN, in which the number of layers and filters was altered according to the user's exigency and particular applications.

In order to ameliorate their model, Ramazi *et al.* [27] used data from wearable sensors, lab tests, and demographic data to create a deep and wide neural network. A hybrid genetic algorithm – extreme learning machine (GA-ELM) algorithm was presented by Alharbi and Alghahtani [28] for the best type 2 diabetes diagnosis. Their model successfully reduced the original dataset to six critical features out of the eight available, resulting in 97.5%.

Kumari *et al.* [29] introduced a framework based on deep learning that combines CNN with LSTM to enhance diabetes prediction utilizing the PIMA-IDD dataset. Through the integration of spatial feature extraction and temporal learning, the model attained improved prediction accuracy when juxtaposed with traditional machine learning and independent deep learning methods.

Using the PIMA-IDD dataset and five-fold and ten-fold cross-validation, Ayon and Islam [30] suggested a deep neural network-based method for diabetes diagnosis. The outcomes signify deep learning efficacy in diabetes prediction with high fidelity and dependability. Nagaraj and Deepalakshmi [31] employed an enhanced support vector machine and a deep neural network for diabetes prediction and screening. The proposed model achieves superior efficacy by employing a deep neural network that utilizes the output of an upgraded support vector machine. The dataset consists of 768 patient records featuring eight key attributes and a target column indicating either a positive or negative result.

The above-mentioned research emphasizes the comparison between different strategies and well-established models, as well as the application of machine learning and deep learning algorithms for diabetes prediction. Nevertheless, the majority prediction algorithms are only able to determine whether the disease is present or absent or calculate the probability of getting it in the future. In order to overcome this drawback, this study suggests a peculiar DNN-based model for advanced diabetes risk prediction.

3. METHOD

3.1. Dataset Description

The National Institute of Diabetes and Digestive and Kidney Diseases offered the PIMA-IDD dataset, which includes 768 medical records, which is used in identifying early risk prediction. Along with a binary target variable (outcome), where 1 denotes diabetes and 0 denotes no diabetes, it comprises an eight clinical parameters, including blood pressure, age, insulin, body mass index (BMI), and glucose levels. Out of 768, 268 samples are with diabetes samples and 500 samples for non-diabetic samples, the dataset is somewhat unbalanced. This imbalance in class distribution poses challenges for classification algorithms. Use of advanced evaluation metrics like precision, recall, F1-score, and receiver operating characteristic–area under the curve (ROC-AUC) can be applied for mitigating the effects of imbalance. For training and testing purposes, the dataset is divided into two subsets: 70% are used for training the model to learn patterns and relationships in the data. 30% are reserved for testing the models performance on unseen data, ensuring an unbiased evaluation [32].

3.1.1. Attributes of the dataset
- Pregnancies: number of times, the patient has been pregnant.
- Glucose: the amount of plasma glucose following a two hour oral glucose tolerance test (OGTT).
- Blood pressure: the diastolic blood pressure in (mmHg).
- Skin thickness: a measure of body fat measured in millimeters of the triceps skin fold thickness.
- Insulin: the serum insulin level after two hours (mu U/ml).
- BMI: measured weight in kg and height to determine fat level.
- Diabetes pedigree function: a function that scores likelihood of Diabetes based on family history.
- Age: the age of the patient, stated in years.
- Outcome (target variable): binary indicator of Diabetes (1- positive, 0 - negative).

3.2. CNN
CNN are a subset of deep learning algorithms created especially in processing time-series data, videos, and other structured grid data. CNNs are very good at extracting spatial and temporal properties from complicated datasets because they resembled the human visual cortex.

3.2.1. Key components of a CNN
- Convolutional layers: use Kernels to perform convolution operations in owing to extract key features such as trends, patterns, and dependencies within the sequence. As the filters move across the input data, feature maps are produced that capture significant representations.
- Activation functions: activation functions like ReLU (Rectified Linear Unit), which give the model non-linearity and enable it to learn complex correlation, are applied after convolution.
- Pooling layers: decrease the spatial dimensions of feature maps, pooling layers like max pooling and average pooling lower computing complexity while preserving important information.
- Fully connected layers: owing to create predictions, these layers connect all of the neurons and flatten the feature maps into a vector. The CNN architecture usually has fully connected layers at the end.
- Dropout layers: by randomly deactivating specific neurons during training, dropout helps to improve generalization and avoid overfitting [33].

3.3. BiLSTM
An improved version of the LSTM network, BiLSTM is made to efficiently capture dependencies in sequential data. BiLSTMs analyze sequences both forward and backward, in contrast to typical LSTMs, which process information in a unidirectional fashion (past to future). Because of its dual viewpoint, the network can take advantage of both past and future information, which makes it especially well-suited for tasks that need for thorough sequence knowledge. BiLSTMs are outfitted with LSTM cells that control the information flow via input, forget, and output gates, guaranteeing that only pertinent features are kept and unnecessary data is deleted. BiLSTMs excel in applications like speech recognition, natural language processing, and medical data analysis by leveraging bidirectional context to model intricate sequential relationships, improving tasks such as diabetes prediction through enhanced temporal pattern analysis [33].

3.4. Attention mechanism with context
A key idea in deep learning is the attention mechanism, which grants models to strenuously concentrate on the most pertinent parts of input data when producing an output. Attention mechanisms were primarily used in machine translation, but they have spread to other fields like computer vision, natural language processing, and healthcare. Attention mechanisms enable models to selectively prioritize important information, increasing accuracy and efficiency, by computing weighted relevance scores for input features.
- Global context: it describes simplified information from the full input, like the aggregated features or general structure.
- External context: describes extra data that is not immediately included in the input, like historical trends, metadata, or past knowledge. For instance, historical trends might serve as an external context in time

series analysis to enhance forecasting. Models can obtain a more thorough comprehension of the input by including these types of context into the attention process, which enables them to:
- Capture intricate feature dependencies.
- Focus more efficiently on important aspects of the data.
- Adapt flexibly to changes in the patterns of input [34].

3.5. Proposed architecture

The section elaborates the proposed deep learning methodology for diabetes prediction. The architecture representation shown in Figure 1 was designed to effectively extract, refine, and classify features from medical data, leveraging the strengths of convolutional operations, bidirectional LSTMs, and attention mechanisms.

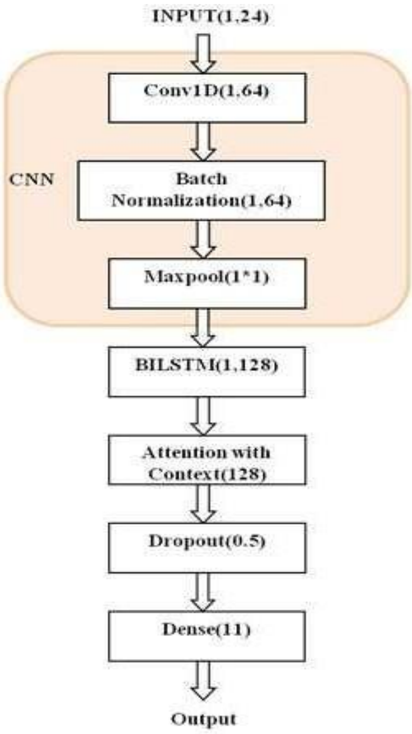

Figure 1. Proposed MLDDAM model's architecture

3.5.1. MLDDAM architecture

- Input: the model's input consists of a feature vector of size (1,24) representing the 24 collected medical attributes relevant to diabetes prediction. These attributes could include various medical parameters, physiological measurements, or other relevant factors related to the prediction task, like glucose levels, age, BMI, blood pressure, or other health indicators. Each feature is normalized to ensure uniform scaling and prevent bias in the learning process.
- Convolutional feature extraction: the first layer of the architecture is a 1D convolutional layer with 64 filters. This layer is designed to capture essential features while preserving the temporal or sequential relationships inherent in the input data. A ReLU activation function is applied to introduce non-linearity and ensure efficient learning. In order to stabilize and hasten convergence, Batch Normalization is exercised following the convolutional operation. Batch Normalization normalizes the activations of the previous layer, lessening internal covariate shifts as well as facilitating faster training.
- Dimensionality reduction: a MaxPooling layer with a 1x1 filter follows the convolutional layer. This pooling operation lessens the feature map's dimension and retains the pivotal factor of computational efficiency while minimizing.
- Temporal feature learning with BiLSTM: to capture temporal dependencies and bidirectional relationships in the sequential input, a BiLSTM layer is employed. The BiLSTM generates an output size of (1,128), effectively summarizing both past and future context within the sequence. This step is critical in understanding the interdependencies among features.

- Attention mechanism with context: to further refine the extracted temporal features, an attention mechanism with context is integrated. This mechanism assigns dynamic weights to the BiLSTM outputs, emphasizing the most relevant features while suppressing irrelevant ones. The attention layer outputs a feature vector of size 128, representing the contextually important information for prediction.
- Regularization with dropout: in order to halt overfitting and boost generalization, a dropout layer with a dropout rate of 0.5 is added, randomly deactivating 50% of the neurons. This stops the model from relying too heavily on specific neurons.
- Dense layers for classification: the refined features then pass across two dense layers. The first has of 11 neurons with a ReLU activation function, which learns intermediate feature representations. The second dense layer has a single neuron with a Sigmoid activation function for binary classification, predicting diabetes presence (1) or absence (0).

3.6. Model training and optimization

The model has 84,673 trainable parameter reflecting its balanced complexity and efficiency. It is trained with a fixed learning rate of 0.0001, chosen to ensure gradual convergence and avoid overshooting the optimal solution. An Adam optimizer is employed for adaptive learning rate adjustments, ensuring robust and efficient training. Binary cross-entropy (BCE) measures the error between the predicted probabilities and the actual class labels.

4. RESULTS AND DISCUSSION

4.1. Performance metrics

The metrics like accuracy, recall, precision, and F1-score to evaluate the models performance.

4.1.1. Accuracy

Accuracy is the proportion of correctly classified instances out of total number of instances, showcasing overall performance but potentially missing class balance in imbalanced datasets. It is illustrated in (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where TP: true positives; TN: true negatives; FP: false positives; FN: false negatives.

4.1.2. Precision

Precision is defined as the percentage of accurately anticipated positive instance among all positive instances. It is illustrated in (2):

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

4.1.3. Recall (sensitivity)

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.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

4.1.4. 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.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

4.2. Experimental results

The proposed model was evaluated using three optimizers-Adam, SGD, and RMSprop at learning rates of 0.01, 0.001 and 0.0001. Performance was measured using Accuracy, Precision, Recall, and F1-Score, as summarized in Table1. Figure 2 illustrates the graphical comparison of various optimizers with a learning rate of 0.01. A comparative evaluation of the proposed MLDDAM model against existing machine learning and deep learning approaches in the related work is summarized in Table 2. All deep learning models were trained for 100 epochs under identical experimental settings to ensure a fair comparison.

Table 1. Performance comparison of different optimizers with different learning rates

Optimizer	Learning rate	Accuracy	Precision	Recall	F1-score
	0.0001	0.9779	0.9717	0.9687	0.9797
Adam	0.001	0.9799	0.9838	0.9732	0.9831
	0.01	0.9899	0.9980	0.9896	0.9837
	0.0001	0.7496	0.7002	0.5223	0.4974
RMSprop	0.001	0.7928	0.6402	0.5090	0.4198
	0.01	0.7214	0.7860	0.6800	0.7881
	0.0001	0.8827	0.8720	0.9712	0.9716
SGD	0.001	0.9389	0.9127	0.9575	0.9695
	0.01	0.8225	0.8396	0.7018	0.8525

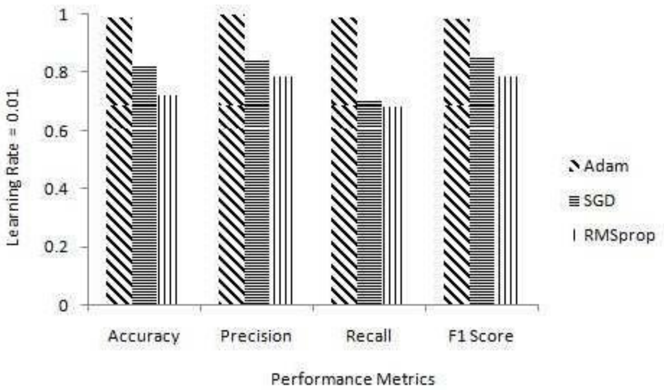

Figure 2. Comparison of different optimizers for learning rate 0.01

Table 2. Quantitative analysis of different existing and the proposed MLDDAM model

Model	Accuracy	Precision	Recall	F1-score
CNN + LSTM [29]	0.9912	0.9763	0.9896	0.9753
DNN [30]	0.9835	-	0.9739	0.9800
SVM + DNN [31]	0.9845	-	0.9943	-
BiLSTM	0.9707	0.9733	0.9716	0.9741
CNN	0.9607	0.9614	0.9661	0.9677
CNN + BiLSTM	0.9883	0.9821	0.9756	0.9795
MLDDAM (Proposed)	0.9943	0.9957	0.9858	0.9907

4.3. Discussion

According to Table 2, the standalone deep learning models, including DNN, CNN, and BiLSTM, demonstrate improved performance, achieving accuracies between 96% and 98%. Hybrid architectures such as SVM+DNN, CNN+LSTM and CNN+BiLSTM further enhance prediction capability by jointly modeling spatial and temporal dependencies, reaching accuracy from 98% to 99%. In contrast, the proposed MLDDAM model achieves the maximum accuracy of 99.43% at 100 epochs, outperforming all baseline and hybrid models. The incorporation of a context-aware attention mechanism is credited with this improvement, which selectively emphasizes the most relevant features while learning spatial-temporal representations through CNN and BiLSTM components. Unlike earlier approaches that rely solely on feature extraction, MLDDAM explicitly prioritizes influential patterns, leading to superior accuracy, recall, precision, and F1-score. Overall, the results confirm that training the MLDDAM architecture for 100 epochs enables more stable convergence and stronger generalization, establishing it as a robust and comprehensive framework for diabetes prediction compared to existing state-of-the-art methods.

5. CONCLUSION

In conclusion, the integration of CNN and BiLSTM with an attention mechanism has proven to be a highly effective approach for determining the presence of diabetes. This hybrid architecture leverages the feature extraction capabilities of CNN, the sequential pattern recognition strength of BiLSTM, and the enhanced focus provided by the attention mechanism. Among the various optimizers explored, the Adam optimizer demonstrated superior performance, effectively fine-tuning the model to achieve optimal results. These findings highlight the robustness and reliability of this approach, offering a promising solution for accurate and efficient diabetes detection.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
J. Jannathul Firthous	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
G. Murugeswari		✓		✓		✓		✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] F. Allam, Z. Nossai, H. Gomma, I. Ibrahim, and M. Abdelsalam, "A recurrent neural network approach for predicting glucose concentration in type-1 diabetic patients," *IFIP Advances in Information and Communication Technology*, vol. 363 AICT, no. PART 1, pp. 254–259, 2011, doi: 10.1007/978-3-642-23957-1_29.

[2] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *Journal of Big Data*, vol. 6, no. 1, p. 13, Dec. 2019, doi: 10.1186/s40537-019-0175-6.

[3] A. Ashiquzzaman *et al.*, "Reduction of overfitting in diabetes prediction using deep learning neural network," in *Lecture Notes in Electrical Engineering*, vol. 449, 2018, pp. 35–43. doi: 10.1007/978-981-10-6451-7_5.

[4] N. Meeral, S. Nisha, and M. Sathik, "Attention mechanism in segmentation and classification of optical coherence tomography images," *Webology*, vol. 18, no. 6, pp. 4859–4873, 2021.

[5] S. Shajun Nisha and M. Nagoor Meeral, "Applications of deep learning in biomedical engineering," in *Handbook of Deep Learning in Biomedical Engineering*, Elsevier, 2021, pp. 245–270. doi: 10.1016/B978-0-12-823014-5.00008-9.

[6] M. Maniruzzaman *et al.*, "Accurate diabetes risk stratification using machine learning: role of missing value and outliers," *Journal of Medical Systems*, vol. 42, no. 5, p. 92, May 2018, doi: 10.1007/s10916-018-0940-7.

[7] D. M. Hawkins and G. J. McLachlan, "Discriminant analysis and statistical pattern recognition," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 695–696, Jun. 1993, doi: 10.2307/2290355.

[8] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326–334, Jun. 1965, doi: 10.1109/PGEC.1965.264137.

[9] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so naive Bayes: aggregating one-dependence estimators," *Machine Learning*, vol. 58, no. 1, pp. 5–24, 2005, doi: 10.1007/s10994-005-4258-6.

[10] S. Brahim-Belhouari and A. Bermak, "Gaussian process for nonstationary time series prediction," *Computational Statistics and Data Analysis*, vol. 47, no. 4, pp. 705–712, 2004, doi: 10.1016/j.csda.2004.02.006.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[12] A. Reinhardt and T. Hubbard, "Using neural networks for prediction of the subcellular location of proteins," *Nucleic Acids Research*, vol. 26, no. 9, pp. 2230–2236, 1998, doi: 10.1093/nar/26.9.2230.

[13] B. Kégl, "The return of ADABOOST.MH: multi-class Hamming trees," *arXiv preprint: 1312.6086*, 2014.

[14] B. P. Tabaei and W. H. Herman, "A multivariate logistic regression equation to screen for diabetes: development and validation," *Diabetes Care*, vol. 25, no. 11, pp. 1999–2003, 2002, doi: 10.2337/diacare.25.11.1999.

[15] I. Jenhani, N. Ben Amor, and Z. Elouedi, "Decision trees as possibilistic classifiers," *International Journal of Approximate Reasoning*, vol. 48, no. 3, pp. 784–807, Aug. 2008, doi: 10.1016/j.ijar.2007.12.002.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[17] P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2019, pp. 367–371. doi: 10.1109/ICCMC.2019.8819841.

[18] S. Wei, X. Zhao, and C. Miao, "A comprehensive exploration to the machine learning techniques for diabetes identification," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, IEEE, Feb. 2018, pp. 291–295. doi: 10.1109/WF-IoT.2018.8355130.

[19] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 105968, Apr. 2021, doi: 10.1016/j.cmpb.2021.105968.
[20] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/j.icte.2021.02.004.
[21] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clinical Epidemiology and Global Health*, vol. 7, no. 4, pp. 530–535, Dec. 2019, doi: 10.1016/j.cegh.2018.12.004.
[22] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Express*, vol. 4, no. 4, pp. 243–246, Dec. 2018, doi: 10.1016/j.icte.2018.10.005.
[23] S. Naveena and A. Bharathi, "A new design of diabetes detection and glucose level prediction using moth flame-based crow search deep learning," *Biomedical Signal Processing and Control*, vol. 77, p. 103748, Aug. 2022, doi: 10.1016/j.bspc.2022.103748.
[24] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction," *CEUR Workshop Proceedings*, vol. 2148, pp. 64–78, 2018.
[25] M. Kowsher, M. Y. Turaba, T. Sajed, and M. M. Mahabubur Rahman, "Prognosis and treatment prediction of type-2 diabetes using deep neural network and machine learning classifiers," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2019, pp. 1–6. doi: 10.1109/ICCIT48885.2019.9038574.
[26] Soniya, S. Paul, and L. Singh, "Application and need-based architecture design of deep neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 13, p. 2052014, Dec. 2020, doi: 10.1142/S021800142052014X.
[27] R. Ramazi, C. Perndorfer, E. Soriano, J.-P. Laurenceau, and R. Beheshti, "Multi-modal predictive models of diabetes progression," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, New York, NY, USA: ACM, Sep. 2019, pp. 253–258. doi: 10.1145/3307339.3342177.
[28] A. Alharbi and M. Alghahtani, "Using genetic algorithm and ELM neural networks for feature extraction and classification of type 2-diabetes mellitus," *Applied Artificial Intelligence*, vol. 33, no. 4, pp. 311–328, Mar. 2019, doi: 10.1080/08839514.2018.1560545.
[29] G. L. A. Kumari, P. Padmaja, and J. G. Suma, "A novel method for prediction of diabetes mellitus using deep convolutional neural network and long short-term memory," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 404–413, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp404-413.
[30] S. Islam Ayon and M. M. Islam, "Diabetes prediction: a deep learning approach," *International Journal of Information Engineering and Electronic Business*, vol. 11, no. 2, pp. 21–27, Mar. 2019, doi: 10.5815/ijieeb.2019.02.03.
[31] P. Nagaraj and P. Deepalakshmi, "Diabetes prediction using enhanced SVM and deep neural network learning techniques," *International Journal of Healthcare Information Systems and Informatics*, vol. 16, no. 4, pp. 1–20, Sep. 2021, doi: 10.4018/IJHISI.20211001.oa25.
[32] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in *Proceedings - Annual Symposium on Computer Applications in Medical Care*, 1988, pp. 261–265.
[33] S. Shajun Nisha, M. Mohamed Sathik, and M. Nagoor Meeral, "Application, algorithm, tools directly related to deep learning," in *Handbook of Deep Learning in Biomedical Engineering*, Elsevier, 2021, pp. 61–84. doi: 10.1016/B978-0-12-823014-5.00007-7.
[34] A. Vaswani *et al.*, "Attention Is All You Need," in *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 6000–6010.

BIOGRAPHIES OF AUTHORS

J. Jannathul Firthous is working as assistant professor at Department of Computer Science, Sadakathullah Appa College, Tirunelveli. She has completed M.Sc., M.Phil in Manonmaniam Sundaranar University, and also qualified in national eligibility test. She is a research scholar of computer science in Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India. She can be contacted at email: janfirth1@gmail.com.

Dr. G. Murugeswari received bachelor of engineering (computer science) from Government College of Engineering, Tirunelveli in 1995 and Master of Information Technology and Computer Engineering from Manonmaniam Sundaranar University, Tirunelveli. She completed her doctorate in computer science and information technology in the year 2017. She is currently working as Associate Professor in Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India. She has more than 15 years of teaching and research experience. Her area of interest includes digital image processing, data mining and bioinformatics. She can be contacted at email: gmurugeswari@msuniv.ac.in.