Advances in Computational Design, *Vol. 11, No. 1 (2026) 1-16*
https://doi.org/10.12989/acd.2026.11.1.016

Implementing LSTM-RNN for improved diabetic dataset classification

Wasim Raja*[1], Chandraprabha K.[2a], V. Ruckmani[3b] and S. Gowri[4c]

[1]Department of Computer Science, Jamal Mohamed College (Autonomous),
(Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India
[2]Department of Information Technology, Bannari Amman Institute of Technology,
Sathyamangalam, Erode, Tamilnadu, India
[3]Department of Computer Science and Applications, KMG College of Arts and Science,
Vellore, Tamil Nadu, India
[4]Department of Computer Applications, Dhanalakshmi Srinivasan College of Arts and Science for Women,
Perambalur, Tamil Nadu, India

(Received July 29, 2025, Revised September 27, 2025, Accepted October 20, 2025)

Abstract. Diabetes is a chronic illness with high morbidity and mortality that influences the quality of life considerably around the globe and thus early and correct prediction is crucial in the effective management and treatment. Nonetheless, the clinical information used is problematic because it is difficult to predict diabetes development in patients, given the complexity and variability of the data. This paper proposes a deep learning-based model with a Long Short-Term Memory (LSTM) recurrent neural network and improved preprocessing and feature selection algorithms. First, Z-score normalization is used to standardize the data, enhancing consistency and identifying abnormalities. Then, to achieve the best feature selection, the Grey Wolf Optimization (GWO) is used to improve predictive performance by identifying the most relevant clinical attributes without falling into local optima. Lastly, the LSTM-RNN model is applied to extract temporal dependencies and latent patterns in the data to correctly classify the data. Through experimentation, it has been shown that the proposed approach clearly exceeds conventional techniques based on all available measures of performance: accuracy; precision; recall; F1 score; and computational efficiency. As indicated by these results, this LSTM-RNN-GWO model shows promise as a valuable resource in the area of predictive analytics related to diabetes care, providing great benefit to patients through its use in early identification of diabetes and subsequent enhancement of their clinical experience.

Keywords: early detection; diabetes condition; GWO; LSTM-RNN; mendeley data; standard deviation; z-score normalization

1. Introduction

The diabetes is a chronic disease in which the pancreas is incapable to supply or secrete enough insulin due to increased blood sugar levels. The incidence of diabetes has increased over the last

---

*Corresponding author, Assistant Professor, E-mail: mwr@jmc.edu
[a] Associate professor, E-mail: chandraprabhak@bitsathy.ac.in
[b] Assistant Professor, E-mail: vruckmani136@gmail.com
[c] Assistant Professor, E-mail: rsgsel@gmail.com

http://www.techno-press.org/acd/5

ISSN: 2092-7614 (Print), 2092-7622 (Online)

few years as of 108 million in 1980 toward 422 million in 2014. The negative belongings of diabetes on body organs consist of the joints, eyes, kidneys, liver, and heart. The diabetes is the leading cause of blindness in the under 50s. Early diagnosis is essential to evade difficulties accompanying through enduring diseases for instance diabetes [1]. Heart rate is the time interval between two consecutive QRS signals on the ECG. RR interval variability is expressed as HRV. A major desirability is that HRV measurements are reproducible and non-invasive. Various Machine Learning (ML) methodologies have been deployed for the automatic diagnosis of diabetes in a non-invasive manner. The Deep Learning (DL) protocols that be able to self-study as of data are now gradually being adopted for diabetes diagnosis [2].

Therefore, the prognosis and initial recognition of diabetes are important for all those who are at risk of developing diabetes. At present, many diseases can be diagnosed using DL methods, which achieve better execution in classification complications. In current years, DL has been handled to diagnose a variety of diseases. DL is a novel investigation way in the field of ML, which has made breakthroughs in recent years in speech acknowledgement and computer vision architectures. DL was first advanced via machine recognition. Body temperature changes, on the other hand, is a widely used technique in the medical field to identify abnormalities. This research has various applications in the medical field, diabetes being no exception. Medical researchers are particularly interested in predicting diseases that are rare compared to the general public. Thus, class heterogeneity is a common problem in most clinical datasets. When there is class inequality, minority groups tend to have significantly lower numbers than other classes. In clinical applications, the misclassification of a minority of patients is costlier than the misclassification of healthy individuals. However, standard learning algorithms always assume homogeneous misclassification errors and symmetric class distributions. It is often difficult to analyze diabetes data to forecast the start of diabetes [3]. To resolve the issue, a Long Short-Term Memory based on Recurrent Neural Network (LSTM-RNN) DL methodology was deployed.

It captures long-term trends and temporal patterns can greatly benefit sequential data processing. This project uses the LSTM-RNN method for diabetes data classification, focusing on improving prediction accuracy and early diagnosis. Using this advanced DL technology, it is designed to effectively learn from time-series data for reliable classification, ultimately improving diabetes management and patient outcomes. LSTM networks solve problems common with traditional RNNs, such as gradient fading and bursting, and are well suited for clinical data analysis where understanding long-term dependencies is important.

1.1 The main contributions of the LSTM-RNN method:

- By capturing the complex patterns and dependencies in the data, the proposed methods produce higher prediction accuracy compared to simple models. This is particularly important for diabetes risk prediction, where accurate, reliable results are critical.
- Diabetes datasets may have missing data or irregularly sampled data. LSTM-CNN can handle these differences better due to its robustness in modeling sequence data, which adapts to real-world scenarios.
- The proposed method is able to automatically extract relevant features from continuous data, thereby reducing the necessity for extensive physical feature engineering. This is specifically useful when handling with complex large-dimension dithering datasets.

2. Literature survey

Pratya Nuankaew et al. [4] discussed that prognostic models in diagnostic systems often use generic factors that may not apply to all populations with different health conditions. To resolve the issue, an Average Weighted Objective Distance (AWOD) methodology was deployed. The proposed methods generally process large volumes of multivariate data to foresee the being of category 2 diabetes. Classification of individual heterogeneity can easily fail if the data set is relatively small.

Suyanto Suyanto et al. [5] discussed that diabetes mellitus is one of the most dangerous and costly diseases. At present, programmed diabetes diagnosis methods are mainly established using DL methodologies, which can classify patients into diabetic and non-diabetic categories with high accuracy. To attain the objective, deployed a Multi-Voter Multi-Commission Nearest Neighbor (MVMCNN) methodology. Unfortunately, the deployed methodology is a very complex and unexplainable black-box model.

Chun Ouyang et al. [6] carried out a useful approach is screening for kind 2 diabetes by means of peripheral pulse waves, can be evaluated quickly, easily, and reasonably with a force transducer at the wrist above the radial artery. A 4-layer multi-task fusion Convolutional Neural Network (CNN) was deployed for feature acknowledgement and trained easily in 30 mins on a used server base.

Sehrish Qummar et al. [7] discussed that diabetes can source vision problems and uniform blindness if not diagnosed premature. To resolve the issue, a Deep Convolution Neural Network (DCNN) methodology was deployed. However, these methods cannot encode basic complex features and can individual classify dissimilar phases of diabetes through very small accurateness, especially in the early stages.

Yunlei Sun et al. [8] discussed that clinical data can achieve reasonable recognition accuracy, and by using readily available and free health record data in daily life, the early diagnosis of diabetes can be made more convenient and faster. However, it is difficult to represent and model the data because of the high dimensionality of data, random errors, incompleteness, systemic biases, and noise heterogeneity.

Supriya Mishra et al. [9] discussed that until now, diabetes testing has been done manually by doctors, which is very time-consuming. To resolve the issue, a DL methodology and its subdivision AI was deployed. DenseNet trains a dataset that automatically detects and accurately classifies diabetes cases in a large dataset. However, the used methodology was taking the larger time to classify the data.

Harshitha et al. [10] discussed that diagnosing abnormalities in the diabetes using another person’s naked eye is time-consuming and expensive, and can lead to incorrect judgments due to subjective differences and considerations between doctors. To resolve the issue, a CNN based DL methodology was deployed. Hence, the proposed methodology leads to the misdiagnosis.

Osama Rabie et al. [11] the author carried out a DL-based Decision Support System (DSS) make use of Bidirectional Long Short-Term Memory (BiLSTM) methodology to exactly diagnosis diabetic sickness as of patient info. Also, DL models are less effective in predicting diabetes due to a lack of appropriate context. Forecasting diabetes as of patient data is considered as a binary label forecast problematic to address the above difficulties.

Ganie et al. [12] discussed that the main goal is to progress a framework on the basis of ML methodologies to predict T2DM via individual lifestyle indicators. A deployed framework based on a ML paradigm using lifestyle indicators is proposed and implemented to better predict T2DM.

To extend the proposed method, it is applied to various large-scale real-time datasets that share common data with T2DM disease to found the performance of the deployed methodology.

Andrés Anaya-Isaza et al. [3] discussed that according to the company, the analysis was conducted using the diabetes routine data set and found that the body temperature patterns of diabetic patients do not follow a particular distribution. To resolve the issue, a DL-based Fourier Transform protocol was deployed. This method can achieve the highest value and complete the classification by CNN ResNet50v2 and can obtain promising results in the problem of warm pattern classification of limited data sets.

Eman AbdelMaksoud et al. [13] carried out an innovative hybrid, DL protocol for diagnosing dissimilar diabetes grades, which named the E-DenseNet methodology. The deployed protocol is a hybrid methodology of the DenseNet and EyeNet on the basis of Transfer Learning (TL). Due to the connection density and convergence of feature maps of all previous layers, the deployed methodology can be very memory intensive, especially as network depth increases.

L. J. Muhammad et al. [14] discussed that diagnostic dataset of diasbetes type 2 was gathered and consumed to progress predictive supervised ML methodologies based on Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Gradient Boosting (GB) methodologies, Naïve Bayes (NB), and K-Nearest Neighbor. Not all complex tasks in ML can be handled, and the computation time for supervised learning is significant.

El-Sayyid et al. [15] discussed that effective DL techniques can automatically detect the disease to diminish the amount of diabetes-interrelated deaths. DL is a novel methodology that encompasses ML technology, a subfield of AI technology that has led to significant progress in clinical applications. Although the performance of the deployed methodology helps clinicians to accurately diagnose diabetes, it takes a long time for classification.

Eman AbdelMaksoud et al. [1] discussed that the Deep CNN (DCNN) architecture is extensively deployed for multi-label classification. It supports to identify ordinary and different grades of diabetes. The author carried out a state-of-art hybrid, DL methodology, which is called E-DenseNet. This is used to classify healthy and unhealthy cells using diabetes datasets. This method has low precision and accuracy for classifying data's so it goes for misdiagnosis.

Ernia Susana et al. [16] discussed that examining methodologies for initial recognition of diabetes are crucial to evade high medical amounts. At present, only offensive surveillance methodologies are commercially accessible. To resolve the issue, ML methodology based on Photoplethysmography (PPG) was deployed. A significant drawback of this approach is patient discomfort while collecting blood samples, while the main limitation of DL is the extensive training segment.

Ordoñez-Guillen et al. [17] discussed that T1DM usually develops at a young age, when an autoimmune reaction stops the pancreas from producing insulin. It is not clear why this problem occurs. It is important to monitor the patient's blood sugar levels because sudden changes in blood sugar levels can be life threatening. To resolve the issue, a novel ML approaches was deployed. The deployed methodologies necessity for excellent data for training, the potential for algorithmic bias, the necessity for skilled experts to develop and maintain machine learning systems, the risk of model overfitting or underfitting, and privacy and data concerns have been raised.

Awab Habib Fakih et al. [18] discussed that timely detection of diabetes prevalence is of great importance to people worldwide. Diabetes is especially a challenge for new hires and employees. Diabetics can be diagnosed if changes in the patient's diet and lifestyle are detected at an early stage. To attain the objective, an Artificial Neural Network (ANN) methodology was deployed. The operation of the deployed methods is unclear, and the resulting solutions do not contain an explanation of how or why they were obtained.

Chaymaa Lahmar et al. [19] evaluated the performance of 28 deep hybrid frameworks for the automatic binary classification of diabetic retinopathy and compare them with 7 end-to-end DL frameworks. By collating DL feature extraction techniques with classic ML techniques, the deployed methodology can classify the target diabetic retina. However, it is time-consuming and data classification requires sophisticated parameter tuning.

Khoula Al Sadi et al. [2] carried out an AI and ML prognostic model will more accurately diagnose T2DM in Oman using a specially developed dataset. The deployed methodology exhibits outstanding execution in pick out the maximum exact analysts of diabetes. Effective training requires large amounts of data, high computing power, and specialized hardware (like GPUs) that are expensive and require technical expertise.

Marmik Shrestha et al. [20] designed to improve accurateness and area further down the curve (AUC) measurements even though dipping processing time for forecasting type 2 diabetes onset. The deployed architecture contains of a DL methodology that uses the SVM protocol beside with the Radial Base Function (RBF) beside with the LSTM for estimate of start of Type 2 Diabetes. This method is not suited for large data sets. If the amount of attributes in every data point is greater than the amount of samples in the training data, the performance degrades.

Nasmin Jiwani et al. [21] discussed that early detection is important to detect and prevent blindness and disfigurement. Manually performing the diagnostic steps for DR by ophthalmologists requires time, effort, and money, increasing the chance of incorrect determination. The scientific public is committed to the development of computer-aided identification systems for initial detection and grading of diabetes strictness. Enduring artificial intelligence investigation highlights the development of DL technology, a mechanism for fine-grained analysis and classification of medical images.

Siddhesh Shelke et al. [22] discussed that unlike computer-aided diagnostic systems, manual diagnosis of diabetes is time-intense and susceptible to to misdiagnosis. To resolve the issue, an InceptionResNet, VGG, and DenseNet methodology was deployed. The main disadvantage is that it avoids integration and initialization, reduces the number of trainable parameters, and reduces accuracy and training time.

3. Materials and Methods

This section illustrates the overall flow of the deployed methodology. Aspects of the developed methodology are shown in Fig. 1. The diabetes dataset in this article was obtained from the Mendeley Data website.

The Z score normalization method is used to use to transform data so that it has a median of 0 and a sd of 1. It further helps in anomaly detection and handling processes. After preprocessed GWO helps improve prediction accuracy of preprocessed dataset by selecting the most relevant features. This is particularly important for clinical data sets where accurate predictions can consume a substantial influence on patient consequences. Then, LSTM-RNN is used to process sequences of data and can capture temporal dependencies and trends. The proposed method is known for its ability to make accurate forecasts in time series analysis.

3.1 Dataset assortment

This article uses the diabetes dataset obtained from the Mendeley Data website. By definition,

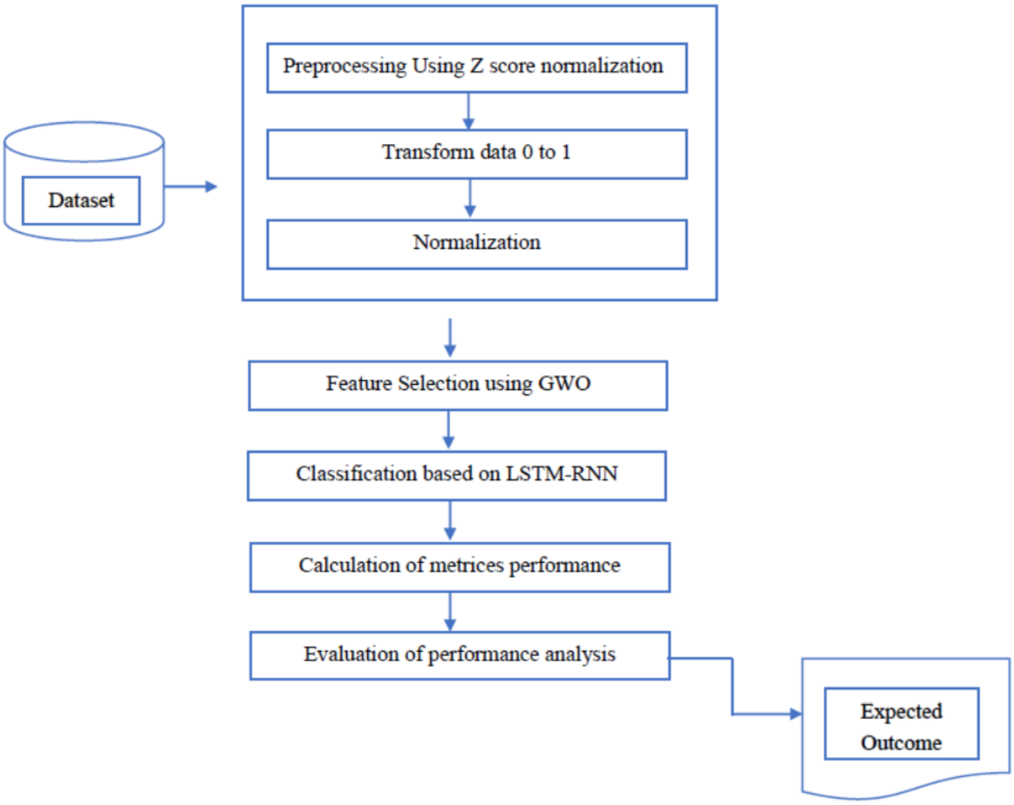

Figure 1. Proposed diagram for LSTM-RNN method

the dataset dropped in this article contains 1000 patient records, and the dataset has 12 different features listed. These characteristics included sex, age, urea, creatinine, HbA1C, cholesterol, triglycerides, HDL, LDL, VLDL, BMI and grade. These target categories indicate whether the patient is prediabetic, diabetic or non-diabetic. Table 1 describes the feature types of the Diabetes dataset.

3.2 Z-score normalization method

In this section we preprocessed the diabetic dataset by using Z-score normalization method. It is also known as standardization, is a preprocessing mechanism deployedto transform data so that it has a median of 0 and a standard deviation of 1. Z-score normalization doesn’t eliminate outliers, but it makes them more noticeable by marking their standard deviation from the mean. It further helps in anomaly detection and handling processes. Normalization is a preprocessing step that warps numerical attribute data so that the values in the data are shifted to a specified range. Generally, several methods are used to normalize the data for instance min-max normalization, Z-score normalization, and decimal normalization. The z-score normalization maps a_i values into an unknown range for attribute C to a'. This can be verified in Eq. (1).

$$a' = \frac{a_i - C_i}{std(C)} \tag{1}$$

Table 1. Details of characteristics types

S.NO	Characteristics	Parameters Description
1	Gender	Male or Female
2	Age	In years, min-max value (20-79)
3	Urea	In mg/dl, min-max value (0.5-38.9)
4	Cr	In µmol/L, min-max value (48,80)
5	HbA1c	In mmol/L min-max value (0.9-16)
6	Chol	In mmol/L min-max value (0.0-10.3)
7	TG	In mmol/L min-max value (0.3-13.8)
8	HDL	In mmol/L min-max value (0.2-9.9)
9	LDL	In mmol/L min-max value (0.3-9.9)
10	VLDL	In mmol/L min-max value (0.1-35)
11	BMI	min-max value (19-47)

where,

a' = outcome of normalization value
a = the assess to be normalized in attribute
C_i = the mean assesses of attribute
$std(C)$ = standard deviation attribute C.

A standard score (often called a Z-score) allows a programmer to calculate the probability that a score occurs within a normal distribution, and compare two scores. This is a very useful statistic because of the different normal distributions. Standardized scores do this by transforming (or standardizing) a normally distributed score from a standard normal distribution into a Z-score.

$$n(xy) = \frac{i(z,y) - \mu}{\alpha} \tag{2}$$

In this,

$n(x,y)$ = New variable
$i(x,y)$ = Odd value
μ = The input value for the accuracy assess of column
α = The input value for the standard deviation of column

$$a' = \frac{a - min_I}{max_I - min_I}(new\_, max_I -) + new\_min_I \tag{3}$$

$$z' = \frac{z - \min(z)}{\max(z) - \min(z)} \tag{4}$$

This iterative technique, also called min-max scaling, is a very simple method. The feature's range is scaled proportionally. Could be [0, 1] or [1, 1]. This has the effect of reducing standard deviation and reducing the significance of outliers. The scope must be determined by considering the characteristics of the data. Below is a general formula to calculate the minimum and maximum values of [0, 1].

$$z' = i + \frac{(z - \min(z))(j - i)}{\max(z) - \min(z)} \tag{5}$$

If z is an original value, then z' is the value after it has been normalized. Preprocessing the diabetes dataset using the Z-score normalization technique effectively normalizes the data by transforming them into a distribution with mean 0 and standard deviation 1. This normalization technique is very useful in DL methodologies because each feature contributes equally to the analysis and can minimize the impact of different scales and units. Overall, the deployed method used on the diabetes dataset to improves the efficacy of the subsequent DL progression, resulting in more reliable and descriptive results.

3.3 Grey Wolf Optimization (GWO) Method

After the preprocess the diabetes dataset is ready to select the features from GWO method. It is an advanced optimization algorithm inspired by the social hierarchy and predator behavior of gray wolves. As it can handle complex multi-dimensional search spaces, it is very useful for feature selection in diabetes datasets. GWO effectively balances exploration (global search) and exploitation (local search) to maximize the likelihood of finding the optimal subset of features. Compared to other optimization methods, this algorithm is simple to implement and requires less parameters. GWO can be easily adapted to different types of data sets and problems, making it a versatile tool for feature selection. Using GWO for feature selection, it can improve the performance of DL models on the diabetic dataset by identifying the most relevant features, thereby reducing dimensionality and improving generalization. Here's a brief overview of how it can be applied:

The hunting progression of GWO is conducted in three phases:

- Chasing, tracking, and process toward the prey
- Harassing, encircling, pursuing the prey
- Assault on the way to the prey

The statistical type of classification of diabetes can be defined in below:

- α is known as fittest solution
- β is known as the second-best solution
- δ is known as third best solution
- ω is known as the balance of candidate solution

The GWO encircling prey can be demonstrated

$$F = |E.A_q(e) - A(e)| \tag{6}$$

$$A(e+1) = A_q(e) - FxD \tag{7}$$

Among them, e indicates the current iteration. F stands for the distance among the gray wolf and its prey. F and E are the coefficient characteristics; A_q is background level character and A is GWO level character. Here, the symbols denote the relative product of every element of two characteristics.

The coefficient characters E and D can be attained by:

$$D = 2d.\ n_1 - d \tag{8}$$

$$E = 2\,x\,n_2 \tag{9}$$

where d is a constant character $(d, d, \ldots, d)$ with constant d decreasing linearly as of 2 to 0 in an iterative progression, and n_1 and n_2 are undefined characters.

The statistical method of encircling prey can be defined: The statistical method of hunting is:

$$F_\alpha = |E_1.A_\alpha(e) - A|F_\beta = |E_2.A_\beta(e) - A|D_\delta = |E_3.A_\delta(e) - A| \tag{10}$$

$$A_1 = A_\alpha - D_1 x F_\alpha A_2 = A_\beta - D_2 x F_\beta A_3 = A_\delta - D_3 x F_\delta \tag{11}$$

By repeating Eq. (6) to Eq. (11), we can finally discovery the optimal solution A_α. The deployed mechanism can be deployed to pick out informative features by slightly modifying Eq. (11) through rounding functions.

3.3.1 Algorithm steps:

Phase 1: Initialize GWO A_x $(x = 1, 2, \ldots, M)$ and characters with max number of iterations f.
Phase 2: Choose α, β and δ wolves from the characteristics.
Phase 3: Update the location of individually wolf by revised Eq. (11).
Phase 4: Calculate Eq. (6) to (11) and go to phase 2.
Phase 5: Print α until the max amount of iterations or convergence of the methodology is met.

The GWO method provides a robust and effective method for feature selection in diabetes datasets, reducing the prediction costs while increasing the performance and interpretability of predictive models. The ability to identify an optimal subset of features makes this a valuable tool in developing advanced diabetes diagnosis and prognostic systems.

3.4 Long Short Time Memory-Recurrent Neural Network (LSTM-RNN)

In this section the diabetes dataset is going to classify based on LSTM-RNN method. An LSTM network is a type of RNN that is suitable for time series prediction and sequence classification problems. Using LSTM-RNN for classification of diabetes datasets exploits temporal patterns in the data to improve prediction accuracy. If the dataset contains time-related data, this method converts the dataset into a format appropriate for time-series examination (for example, by serializing historical data). LSTM-RNN is used to process sequences of data and can capture temporal dependencies and trends. The proposed method is known for its ability to make accurate forecasts in time series analysis. For diabetes risk prediction, this means being able to provide more reliable predictions based on past data patterns. There are two main, interrelated reasons to use an LSTM-RNN rather than a single model: Benefits of the LSTM-RNN method.

The deployed methodology is able to make supplementary accurate predictions and achieve higher performance levels than the individual models it provides. This helps to reduce predictive variability and the overall performance of the framework. In this step, we deploy the LSTM-RNN methodology to classify the data as churn data or non-churn data. The regular structure of LSTM solves the problem of the virtual vanishing of gradients and can efficiently transmit valuable data across the LSTM network. An RNN cannot accurately catch the long-term needs between feature characters. The LSTM unit consists of an input gate (gt), a forget gate (yt), and an output gate (zt). This threshold controls the activation vector of the memory cell.

3.4.1 Forget Gate (yt)

By use of the present input a_t and the hidden layer st_{-1}, the "forget gate" decides how many info from the previous layer pt_1 should be forgotten and how much information should be retained. The formula for a forgetting gate is given by Eq. (12).

$$y_t = Sigmoid(v_{aya_t} + v_{sy}s_{t-1} + k_z) \tag{12}$$

meanwhile, k_z , v_{ay}, and v_{sy} represent the bias vectors of the weighting matrix between a_t and y. The weighted matrix is in l that corresponds to s_{t-1}.

3.4.2 Input Gate (gt)

This can be used in Eq. (13) to determine how many of the network's inputs a_t should remain in the current cell state s_t.

$$g_t = Sigmoid(v_{aga_t} + v_{sg}s_{t-1} + k_g) \tag{13}$$

here k_z @comm v_{sg} denotes the bias vector, the weighted matrix is between a_g and g_t, and the weighted matrix is correspondingly between s_{t-1} and g_t.

3.4.3 Output Gate (zt)

The current output s_t can be used to calculate the total number of inputs to the LSTM at cell state p_t. LSTM gates are fully-connected networks with vector inputs and real outputs. Eq. (14) can be used to describe output gates.

$$z_t = Sigmoid(v_{aza_t} + v_{sz}s_{t-1} + k_z) \tag{14}$$

k_z, v_{az}, who denote the bias vectors, while the weight matrix is between a_t and 0, and the weight matrix is also between s_{t-1} and zt.

The final output of the LSTM cell consists of the cell output state p_t and the layer output s_t, defined by Eqs. (15) and (16), respectively.

$$P_t = y_t \Theta P_{t-1} + g_t \Theta \breve{P}_t \tag{15}$$

$$s_t = z_t \Theta tans(P_t) \tag{16}$$

The intermediate cell input is denoted as p_t that is expressed in Eq. (17):

$$\breve{P}_t = tans(V_{ap}a_t + V_{sp}s_{t-1} + k_p) \tag{17}$$

here, the bias vectors are represented by k_z, v_{ap}, and v_s, and the weight matrices are between w_g and p_t and the weight matrices between s_{t-1} z_a and p_t. The LSTM-RNN is illustrated in the Fig. 2.

LSTM-RNN has been used in various fields to predict diabetes from patient datasets containing 1000 features. It is a combination of two different data from two directions. Maximum and minimum values obtained from patient data were then combined. Despite the distributed word embeddings, the forward and backward variables of the LSTM-RNN are independent of each other. Finally, the hidden layers of the LSTM-RNN are determined in Eq. (18) to combine the forward and backward vectors during time step t.

$$s_t = [\vec{s}_t, \overleftarrow{s}_t] \tag{18}$$

In addition, DL technology improves the classification accuracy. However, layered LSTM-RNN is more effective for shallow learning techniques. Therefore, the proposed method determines the layered LSTM-RNN using the local environment and symmetrical complex data. In LSTM, the lower layer output serves as the upper RNN input. The LSTM-RNN method has

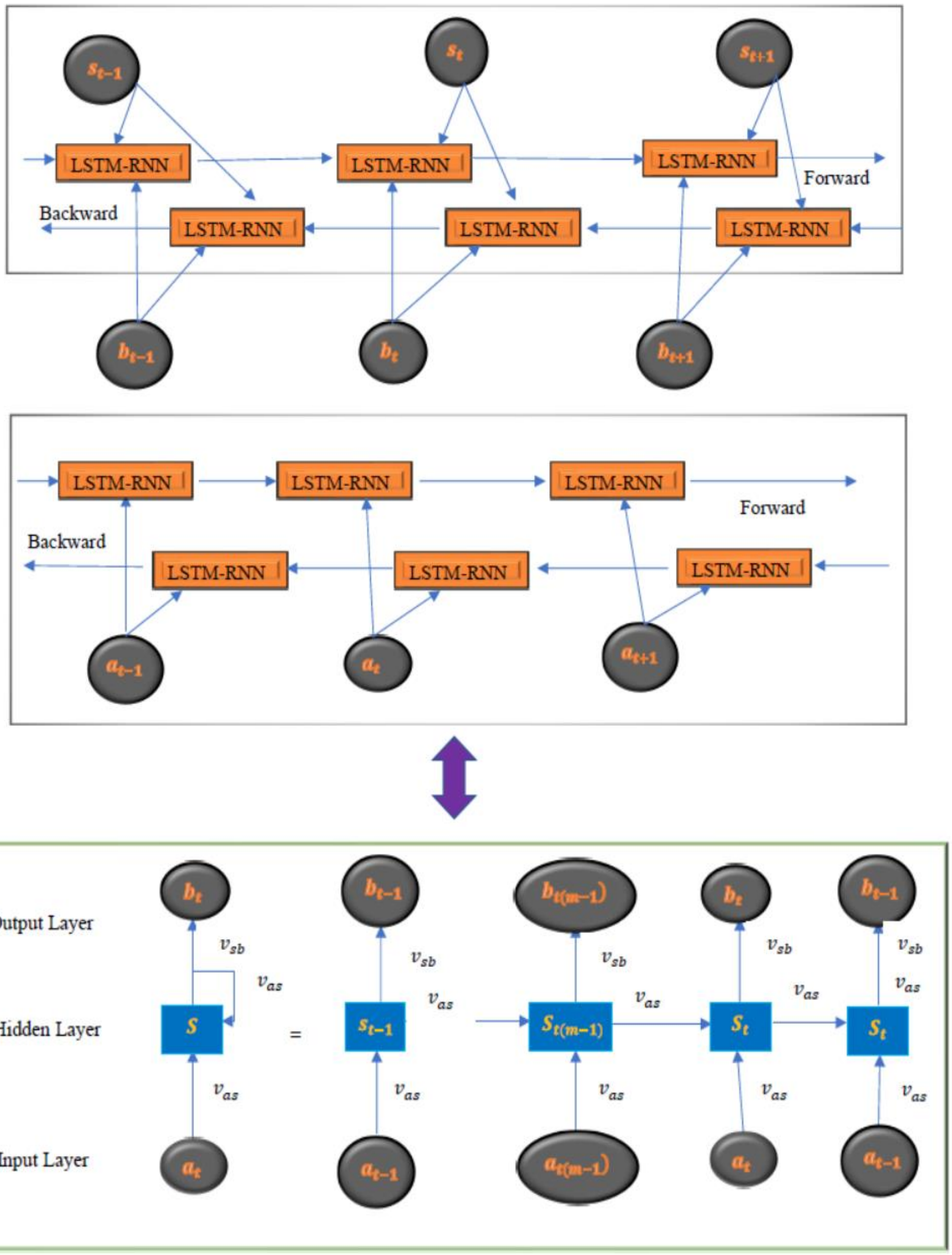

Figure 2. Architecture diagram of LSTM-RNN method

Table 2. Simulation parameters

Parameters	Values
Dataset Name	Diabetes Dataset
Number of Records in Dataset	1000
Used Language	Python
Used Tool	Anaconda

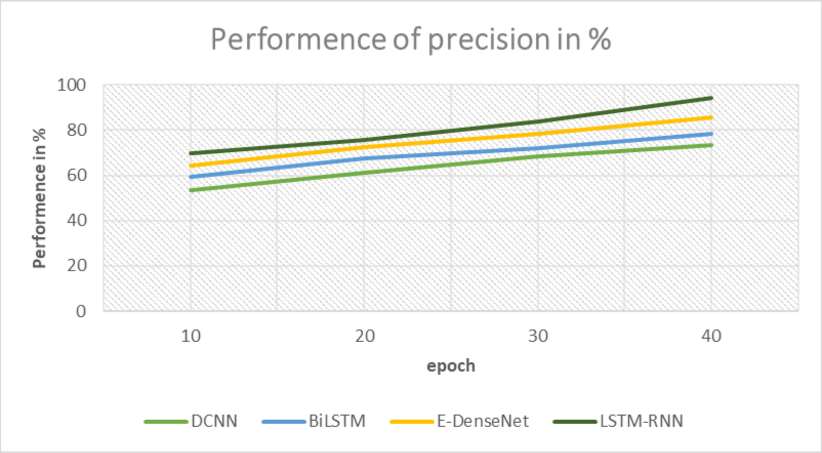

Figure 3. Performance analysis of precision in %

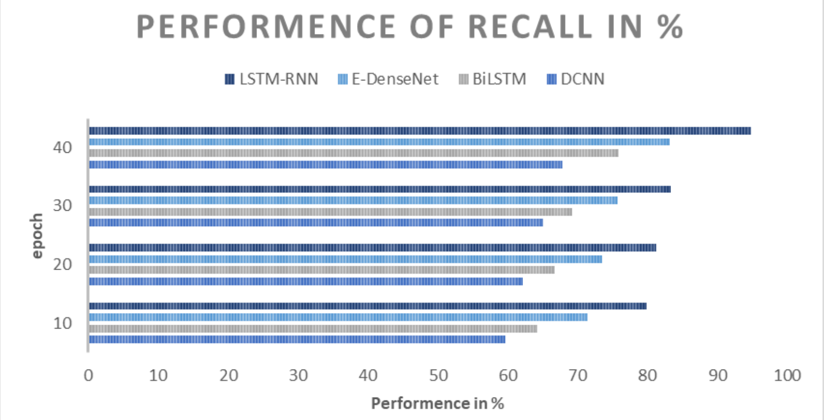

Figure 4. Performance analysis of recall in %

been proven to be a powerful tool for diabetes classification using diabetes datasets, allowing for more accurate and efficient processing of continuous data. Despite the challenges of computational complexity and preprocessing requirements, its advantages make it a valuable approach in the fields of clinical data analysis and disease prognosis. Further research and optimization could open up more possibilities for clinical applications.

4. Result and discussions

The performance of the developed method is estimated via precision, recall, accuracy, F1 score and time complexity. In this evaluation, the existing DCNN, BiLSTM, and E-DenseNet methods were used to find a reliable and accurate diabetes detection method through the introduced method. Table 2 shows the simulation results and parameters in this article.

The diabetes dataset is used to classify the patient’s medical history and contains 1000 datasets. The execution method is implemented using Python and Anaconda tool.

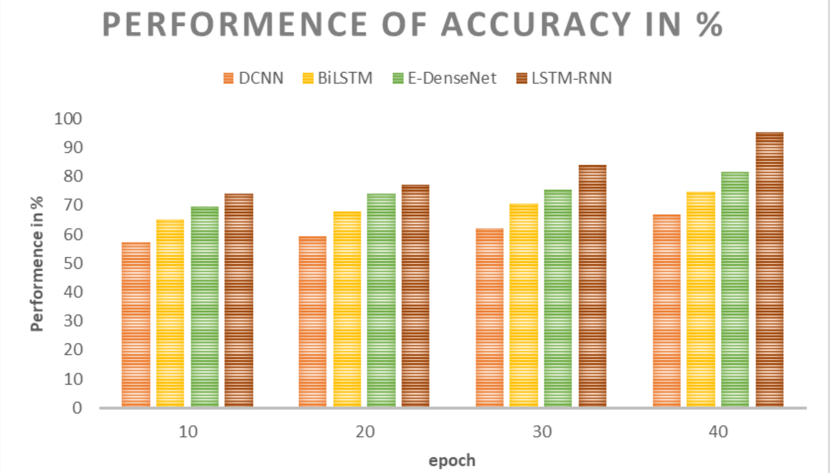

Figure 5. Performance analysis of accuracy in %

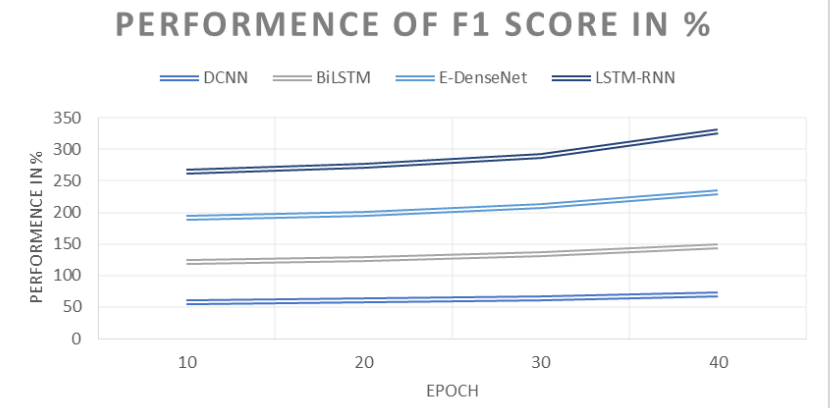

Figure 6. Performance analysis of F1 score in %

Fig. 3 demonstrates that the execution precision of DCNN is 73.54%, BiLSTM is 78.49%, E-DenseNet is 85.63%, and LSTM-RNN is 94.21%. The methodology introduced has improved precision over the previous methodology. High precision increases patient confidence in healthcare provider and model predictions. If the model identifies a case as high-risk, health professionals can be more confident about intervention or additional testing.

Fig. 4 demonstrates that the execution recall of DCNN is 67.87%, BiLSTM is 75.89%, E-DenseNet is 83.14%, and LSTM-RNN is 94.87%. The methodology introduced has improved recall over the previous methodology. The combination of high recall and precision allows for a balanced evaluation of sample functioning. The precision concentrations on the accurateness of optimistic estimates, while recall ensures the model does not miss too many positive events.

Fig. 5 demonstrates that the execution accuracy of DCNN is 66.78%, BiLSTM is 74.71%, E-DenseNet is 81.47%, and LSTM-RNN is 95.21%. The methodology introduced has improved accuracy over the previous methodology. High accuracy reflects a balance between accurately identifying true positives and true negatives and suggests that the model is not biased toward one class over the other.

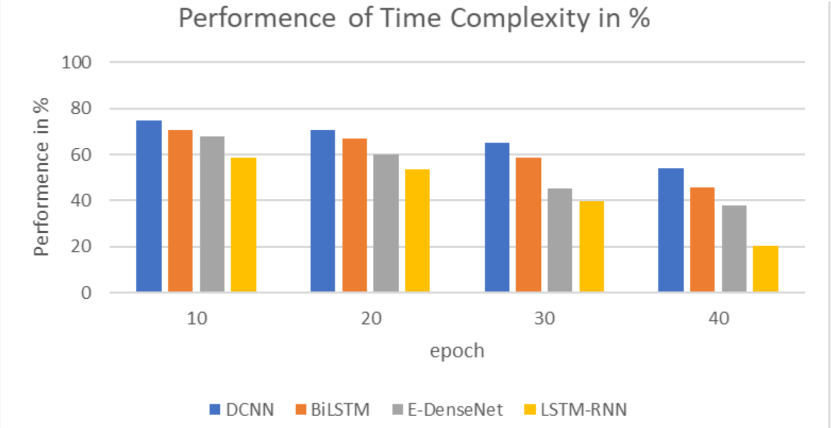

Figure 7. Performance analysis of time complexity in %

Fig. 6 illustrates that the execution F1 score of DCNN is 69.96%, BiLSTM is 76.41%, E-DenseNet is 85.64%, and LSTM-RNN is 95.78%. The methodology introduced has improved accuracy over the previous methodology. By achieving a high F1 score, the model ensures that most patients at risk of diabetes are accurately identified (high recall) and that patients found to be at risk are indeed at risk (high precision). Improve clinical outcomes by ensuring timely, precise intervention.

Fig. 7 shows that the time complexity performance of DCNN is 53.87%, BiLSTM 45.78%, E-DenseNet 37.84%, and LSTM-RNN 20.47%. Due to the introduced method, the time complexity is much less than the previous method. Less time-consuming method can handle large data sets efficiently. As the volume of clinical data increases, the ability to process this data quickly becomes increasingly important.

5. Conclusions

In this novel for diabetic dataset classification the LSTM-RNN method was deployed. Similarly, in this paper, we evaluated the narrative and comprehensiveness of the z-score normalization method using the diabetes dataset. The z-score normalization method is deployed to transform data so that it has a median of 0 and a standard deviation of 1. Z-score normalization doesn't eliminate outliers, but it makes them more noticeable by marking their standard deviation from the mean. After preprocessed the features were selected in the dataset by using the GWO method. It helps improve prediction accuracy of preprocessed dataset by selecting the most relevant features. It then balances exploration (searching for new areas in the feature space) and exploitation (focusing on finding the best solution). After selected the features of preprocessed dataset we go to the classification phase by the use of LSTM-RNN method. The proposed method is able to automatically extract relevant features from continuous data, thereby reducing the necessity for extensive physical feature engineering. This is particularly useful when handling with complex large-dimensional dithering datasets. Furthermore, the LSTM-RNN method achieved well and achieved a precision of 94.21%, recall 94.87%, accuracy 95.21%, F1 score 95.78%, time complexity 20.47% by using the diabetes dataset. LSTM networks solve problems common with traditional RNNs, such as gradient fading and bursting, and are well suited for clinical data analysis where understanding long-term dependencies is important.

References

1. AbdelMaksoud, E., Barakat, S., Elmogy, M. (2022). A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. Medcal & Biological Engineering & Computing, 60(7), 2015-2038. https://doi.org/10.1007/s11517-022-02564-6
2. Al Sadi, K., Balachandran, W. (2023). Prediction model of Type 2 diabetes mellitus for oman prediabetes patients using artificial neural network and six machine learning classifiers. Applied sciences, 13(4), 2344. https://doi.org/10.3390/app13042344.
3. Anaya-Isaza, A., Zequera-Diaz, M. (2022). Fourier transform-based data augmentation in deep learning for diabetic foot thermograph classification. Biocybernetics and Biomedical Engineering, 42(2), 437-452. https://doi.org/10.1016/j.bbe.2022.03.001.
4. Nuankaew, P., Chaising, S., Temdee, P. (2021). Average weighted objective distance-based method for type 2 diabetes prediction. IEEE Access, 9, 137015-137028. https://doi.org/10.1109/ACCESS.2021.3117269.
5. 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.
6. 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.
7. Qummar, S., Khan, F.G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z.U., ... Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection. Ieee Access, 7, 150530-150539. https://doi.org/10.1109/ACCESS.2019.2947484.
8. Sun, Y., Zhang, D. (2019). Diagnosis and analysis of diabetic retinopathy based on electronic health records. Ieee Access, 7, 86115-86120. India, October.
9. Mishra, S., Hanchate, S., Saquib, Z. (2020October). Diabetic retinopathy detection using deep learning. In 2020 International conference on smart technologies in computing, electrical and electronics (ICSTCEE). IEEE. India, October.
10. Harshitha, C., Asha, A., Pushkala, J.L.S., Anogini, R.N.S. (2021). Predicting the stages of diabetic retinopathy using deep learning. In 2021 6th international conference on inventive computation technologies (ICICT). IEEE. India, January.
11. 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. https://doi.org/10.3389/fpubh.2022.861062.
12. Ganie, S.M., Malik, M.B., Arif, T. (2022). Performance analysis and prediction of type 2 diabetes mellitus based on lifestyle data using machine learning approaches. Journal of Diabetes & Metabolic Disorders, 21(1), 339-352. https://doi.org/10.1007/s40200-022-00981-w.
13. AbdelMaksoud, E., Barakat, S., Elmogy, M. (2020). Diabetic retinopathy grading based on a hybrid deep learning model. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI). IEEE. Bahrain, October.
14. Muhammad, L.J., Algehyne, E.A., Usman, S.S. (2020). Predictive supervised machine learning models for diabetes mellitus. SN Computer Science, 1(5), 240. https://doi.org/10.1007/s42979-020-00250-8.
15. El-Bashbishy, A.E.S., El-Bakry, H.M. (2024). Pediatric diabetes prediction using deep learning. Scientific Reports, 14(1), 4206. https://doi.org/10.1038/s41598-024-51438-4.
16. Susana, E., Ramli, K., Murfi, H., Apriantoro, N.H. (2022). Non-invasive classification of blood glucose level for early detection diabetes based on photoplethysmography signal. Information, 13(2),

59. https://doi.org/10.3390/info13020059.

17. Ordoñez-Guillen, N.E., Gonzalez-Compean, J.L., Lopez-Arevalo, I., Contreras-Murillo, M., Aldana-Bobadilla, E. (2023). Machine learning based study for the classification of Type 2 diabetes mellitus subtypes. BioData Mining, 16(1), 24. https://doi.org/10.1186/s13040-023-00340-2.
18. Fakih, A.H., Venkatesh, A.N., Vani, Naved, M., Kshirsagar, P.R., Vijayakumar, P. (2022). An efficient prediction of diabetes using artificial neural networks. In AIP Conference Proceedings,2393(1), 020071. AIP Publishing LLC. May.
19. Lahmar, C., Idri, A. (2023). Deep hybrid architectures for diabetic retinopathy classification. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(2), 166-184. https://doi.org/10.1080/21681163.2022.2060864.
20. 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. https://doi.org/10.1007/s11042-022-13582-9.
21. Jiwani, N., Gupta, K., Afreen, N. (2022). A convolutional neural network approach for diabetic retinopathy classification. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT). India, April.
22. Shelke, S., Subasi, A. (2023). Detection and classification of Diabetic Retinopathy Lesions using deep learning. In Applications of artificial intelligence in medical imaging. Academic Press, Cambridge, Massachusetts, USA.
23. 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
24. Sun, Y., Zhang, D. (2019). Diagnosis and analysis of diabetic retinopathy based on electronic health records. Ieee Access, 7, 86115-86120 https://doi.org/10.1109/ACCESS.2019.2918625
25. 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), 148. https://doi.org/10.1186/s13638-020-01765-7.