AIMS Bioengineering, 13(2): 261–280.
DOI: 10.3934/bioeng.2026012
Received: 10 April 2026
Revised: 28 May 2026
Accepted: 03 June 2026
Published: 22 June 2026

Bioengineering

https://www.aimspress.com/journal/Bioengineering

Review

Diabetes prediction using a context-adaptive activation neural network

Ren Chen[1], **Chao Wang**[2] **and Fulong Chen**[2,*]

[1] School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui Province 230026, China

[2] School of Computer and Information, Anhui Normal University, Wuhu, Anhui Province 241002, China

* **Correspondence:** Email: long005@ahnu.edu.cn; Tel: +865535910371.

Abstract: This paper describes a novel diabetes prediction framework underpinned by a neural network architecture which features a parameterized activation function, called the Self-Learning Activation Linear Unit (SLALU). The proposed diabetes prediction model is composed of multiple self-learning activation linear units with parameters, and this SLALU can fit most of the traditional activation functions, including but not limited to the Rectified Linear Unit (ReLU) and the Parametric Rectified Linear Unit (PReLU). This self-adjusting activation approach provides a flexible bending ability, thereby outperforming the bending capabilities of traditional activation functions and reducing the intricacy involved in their manual selection. The design of our neuron model improves the network's representational capacity and fitting ability. Deployed in diabetes prediction, our advanced network model demonstrated superior performance across a suite of metrics, thereby recording an accuracy of 98.18%, thus outperforming several existing models.

Keywords: diabetes prediction; attention mechanism; deep neural networks; adjustable activation neural networks

1. Introduction

Diabetes represents a global health challenge, primarily attributed to disruptions in insulin hormone secretion by the pancreas [1]. Insulin plays a critical role in transforming dietary glucose into energy and facilitating its absorption into the bloodstream. A malfunctioning pancreas will produce insufficient insulin and can precipitate the onset of diabetes, which is a disease related to serious health complications such as renal failure and cardiovascular disease [2]. The incidence of diabetes, especially among adults, has shown an upward trend in recent years, with developing countries bearing a significant brunt of this increase. Projections suggest a staggering rise in diabetes

prevalence, with nearly 48% of the global population expected to be affected by 2045 [3]. This alarming trajectory underscores the substantial impact diabetes has not only on individuals but also on societal health infrastructure, thus emphasizing the need for preemptive strategies in diabetes prevention and management.

Diabetes is mainly classified into two types: Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D). In patients with T1D, the number of pancreatic islet cells decreases or even disappears,thus resulting in insufficient or complete absence of insulin secretion, which, in turn, causes elevated blood glucose levels as the body cannot effectively utilize insulin. In patients with T2D, the ability of insulin to regulate glucose metabolism is impaired (i.e., insulin resistance), accompanied by functional defects of pancreatic islet cells, ultimately leading to elevated blood glucose levels. The application of artificial intelligence algorithms provides a novel and effective approach to meet the needs for early warning and a timely diagnosis of diabetes.

In this study, our focus is on the prediction of T1D. Although current artificial intelligence research has made some progress in the diagnosis of T1D, such as Deep Neural Networks (DNN) [4–9] and Calibrated Multimodal Learning (CML) [10–15], challenges remain in handling data imbalances and improving the model transferability. To address the complexity of the prediction task, we draw inspiration from the area of nonlinear activation functions [16] in neural networks, thereby introducing a neural network with a self-learning activation method featuring parameters.

Nonlinear activation functions empower neural networks to model complex, high-dimensional data. Traditional activation functions such as the Sigmoid function and Rectified Linear Units (ReLUs) [17] have been instrumental in mitigating issues such as vanishing gradients, while further enhancements have led to the development of parameterized activation functions. These parameterized functions, such as Parametric ReLUs(PReLUs) [18], have not only demonstrated superior representational capabilities and faster convergence rates compared to conventional activations, but they also contribute to enhancing the predictive accuracy of models [18–20].

In exploring the impact of parameterized activation functions on improving the model accuracy and reducing the training time, our experiments have indeed identified a significant advantage when predicting diabetes. These benefits are not only reflected in an enhanced predictive accuracy but also in a faster model convergence. Upon verifying the efficacy of parameterized activations in diabetes prediction, we introduced a new method of parameterized activation functions and constructed a Self-Learning Activation Linear Unit (SLALU) for diabetes prediction.

This work integrates diabetes prediction with advances in nonlinear activation functions to improve the predictive accuracy. The key contributions of this paper are as follows:

- We propose a self-adjusting activation approach, which provides a flexible bending ability, thereby outperforming the bending capabilities of traditional activation functions and reducing the intricacy involved in their manual selection.
- We propose a diabetes prediction framework based on SLALU, and experimental results demonstrate an excellent performance of the model in terms of accuracy and convergence speed.

This paper is organized as follows: first, we review the related works on diabetes predictions, then introduce SLALU along with the theoretical foundation it is built upon. subsequently, we showcase the experimental setup and results before concluding in the last section. Our objective is to substantiate the significance and application value of nonlinear activation functions within the field of diabetes

prediction by providing a robust theoretical and empirical basis and thus illuminate future research in this sphere.

2. Related works

Diabetes has been recognized as one of the most serious chronic metabolic diseases worldwide. Data from the International Diabetes Federation (IDF) showed that about 537 million adults were diagnosed with diabetes in 2021. The number is expected to rise to 783 million by 2045 [21]. Early detection and accurate prediction of diabetes are considered critical to prevent complications and reduce medical burdens. With the development of artificial intelligence, machine learning and deep learning are widely used in diabetes predictions and show obvious advantages. Traditional machine learning has been extensively applied. Deep learning represented by multilayer perceptron and convolutional neural networks(CNNs) has been proven to further improve the prediction accuracy and the generalization ability.

This chapter summarizes representative studies on diabetes prediction using machine learning and deep learning published between 2022 and 2025, especially those based on the Pima Indians Diabetes Dataset. The method systems, performance, advantages, limitations, and technical trends were analyzed to provide systematic references for researchers and clinical practitioners.

Before deep learning was widely used, traditional machine learning algorithms were extensively applied in diabetes prediction. Deep learning has become a dominant approach for time-series forecasting in diverse fields, such as stock market analyses [22]. These methods are regarded as important for clinical decision support due to the good interpretability and the relatively low computational cost. Whig et al. [23] employed the low-code PyCaret machine learning framework to automatically construct, train, and compare multiple classifiers, and optimized the model through hyperparameter tuning for the classification, detection, and prediction of diabetes. This approach achieved an accuracy of approximately 90%. However, due to its implementation based on a low-code platform, the interpretability of the model structure and internal mechanism remains relatively poor. AlZu'bi et al. [24] proposed a diabetes monitoring system combined with big data technology, which adopted various methods including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Adaptive Boosting, and deep learning for diabetes prediction, thus achieving an accuracy of 94.6%. However, the system has high requirements for computing resources when operating in a big data environment, with certain limitations in its application. Smith et al. [25] used the ADAP neural network algorithm, and a classification accuracy of 76% was achieved; it was considered groundbreaking for early neural network methods in diabetes prediction. However, a simple model structure was adopted, and the ability to extract complex nonlinear features was limited. Rough data preprocessing and feature engineering were also found. Saxena et al. [26] employed four classifiers. Three feature selection methods were adopted. The Pima Indians Diabetes Dataset was used for classification and prediction.The optimal performance was achieved by the Random Forest classifier, with an accuracy of 79.8%. However, the specificity was found to be unsatisfactory, and the ability to identify healthy individuals needs to be enhanced. Mushtaq et al. [27] used six classifiers: Logistic Regression, SVM, KNN, Gradient Boosting, Naive Bayes, and Random Forest. An accuracy of 80.7% was obtained by Random Forest on the SMOTE-balanced dataset. Accuracies of 82.0% and 81.7% were achieved by the voting ensemble

model on the original and balanced datasets, respectively. However, the overall accuracy was considered relatively low, and the weight assignment of the voting ensemble method was regarded as simple. Chou et al. [28] used four machine learning algorithms: logistic regression, neural network, decision jungle, and decision tree. A comparative study was conducted on diabetes prediction for 15,000 female patients from Taipei City Hospital. Prediction models were built based on eight features. The best performance was achieved by the decision tree model, with an area under the curve (AUC) of 0.991, an accuracy of 95.3%, a precision of 92.7%, a recall of 93.1%, and an F1-score of 92.9%. However, the study was limited to female subjects. The applicability of the model to male patients was not verified. Deep learning methods were not adopted to explore deeper feature representations. With the improvement of computing power and the accumulation of big data, deep learning methods have been widely studied in the field of diabetes prediction. Multilayer Perceptron (MLP), CNN, and their variants have been proven to show significant advantages in processing complex nonlinear medical data. Al Reshan et al. [29] proposed an innovative ensemble deep learning clinical decision support system for diabetes prediction. A stacked ensemble model was constructed by combining artificial neural network (ANN), long short-term memory (LSTM) and CNN, with stack-ANN, stack-LSTM, and stack-CNN used as metamodels. The Extra Tree Classifier was adopted for feature selection. However, the model structure was considered complex with high demand for computing resources, and the interpretability of stack-ANN was found to be relatively weak. Waughfa et al. [30] compared the effectiveness of various machine learning and deep learning techniques, including SVM, MLP, Random Forest, KNN, and Decision Tree, using the Pima Indian dataset. A comprehensive evaluation was performed based on metrics such as accuracy, precision, and recall. MLP was identified as the most effective algorithm, thereby achieving an accuracy of 85%. However, the accuracy remained relatively low, thus indicating that there is still room for improvement in the model performance. Soltanizadeh et al. [31] designed a low-complexity deep learning model for T2D diagnoses. Experiments were conducted on the PIMA Indian Diabetes Dataset, and trade-off analyses between the accuracy and the complexity were performed on several deep learning architectures. The proposed hybrid architecture (CNN+MLP) was found to perform best, with an accuracy of 93.89%. However, validation was only performed on a single dataset, and the low-complexity design may be accompanied by a partial loss of the prediction accuracy.

Ensemble learning methods are used to combine the prediction results of multiple base learners. A better generalization performance can be obtained over using a single learner. In recent years, stacking ensemble, voting ensemble and other methods have been widely studied in diabetes prediction. Ateequr Rahaman et al. [32] applied advanced ensemble methods including Random Forest, Boosting, Bagging and Stacking. The Pima Indians Diabetes Database was balanced with SMOTE for diabetes prediction. Ensemble methods, especially Random Forest, Bagging and Boosting, were proved to be better than traditional models, with an accuracy of 88%, a recall of 82% and a precision of 85%. However, the accuracy of 88% was considered to be improvable, and the further combination of deep learning and ensemble learning was not explored. Zhuang et al. [21] proposed a hybrid prediction framework combining CNNS and ensemble learning. Multiple classifiers were integrated using a soft voting strategy. Evaluation was conducted on the Pima Indians Diabetes Dataset. The proposed CNN-Voting ensemble model was found to perform best, thereby achieving an accuracy of 98%, an F1-score of 0.99, and a recall of 99% on the maximum dataset. However, the model structure was considered complex with high computational resource requirements, and the method for determining classifier weights in

soft voting was regarded as needing further optimization. Andrade-Arenas et al. [33] used SMOTE to balance the Pima Indian dataset. Nine machine learning models were evaluated, with a focus on ensemble methods such as Random Forest and AdaBoost. Random Forest and AdaBoost were found to produce the most robust and consistent results on key metrics including AUC-Receiver Operating Characteristic (ROC) and Area Under the Precision-Recall Curve (AUPRC). However, deep learning models were not included for the comparative analysis, and the interpretability of Random Forest and AdaBoost was considered relatively low.

3. Diabetes prediction and methodology

In this paper, the Pima Indians Diabetes Dataset(PIDD) consists of 768 female diabetes patients from the Pima Indian population near Phoenix, Arizona. The dataset was collected on Kaggle *. It is composed of 268 diabetes-positive instances and 500 non-diabetic instances, as shown in Table 1. We only applied oversampling to the training set (SMOTE). The dataset was randomly split into 80% training and 20% testing using a fixed random seed (42). All experiments were repeated 5 times, and the reported results are averages. We proposed a comprehensive framework for diabetes prediction, which was evaluated using a publicly available PIDD. The experimental results demonstrate that our model achieves a higher accuracy and stability than current models in predicting diabetes. Moreover, we developed a novel diabetes prediction model, based on a personalized learning neural network with a joint attention mechanism. Specifically, we used a fully connected layer to map the input data to a hidden vector and then computed the output weights of each neuron using adaptive activation functions, which can be adjusted during training based on the input data. Next, we introduced a joint attention mechanism to weigh the hidden vector, thus capturing the importance of different features. Then, the attention weight features are linearly fused with the high-dimensional features output by the model. Finally, a softmax layer was used to output the prediction result. The diabetes prediction framework is shown in Figure 1.

Table 1. Pima Indians diabetes dataset.

Attributes	Description
Pregnancies	This feature represents the number of pregnancies a person has had.
Glucose	This feature indicates the glucose level in the blood, which is an important indicator of diabetes.
BloodPressure	This feature represents the measurement of blood pressure.
SkinThickness	This feature denotes the thickness of the skin, which can also be relevant in diabetes prediction.
Insulin	This feature indicates the insulin level in the blood, which is closely associated with diabetes.
BMI	This feature represents the body mass index, which is a measure of body fat based on height and weight.
Pedigree	This feature represents the percentage of diabetes prevalence in the family pedigree.
Age	This feature indicates the age of the individual.
Outcome	This is the final result of the prediction.

*http://www.kaggle.com

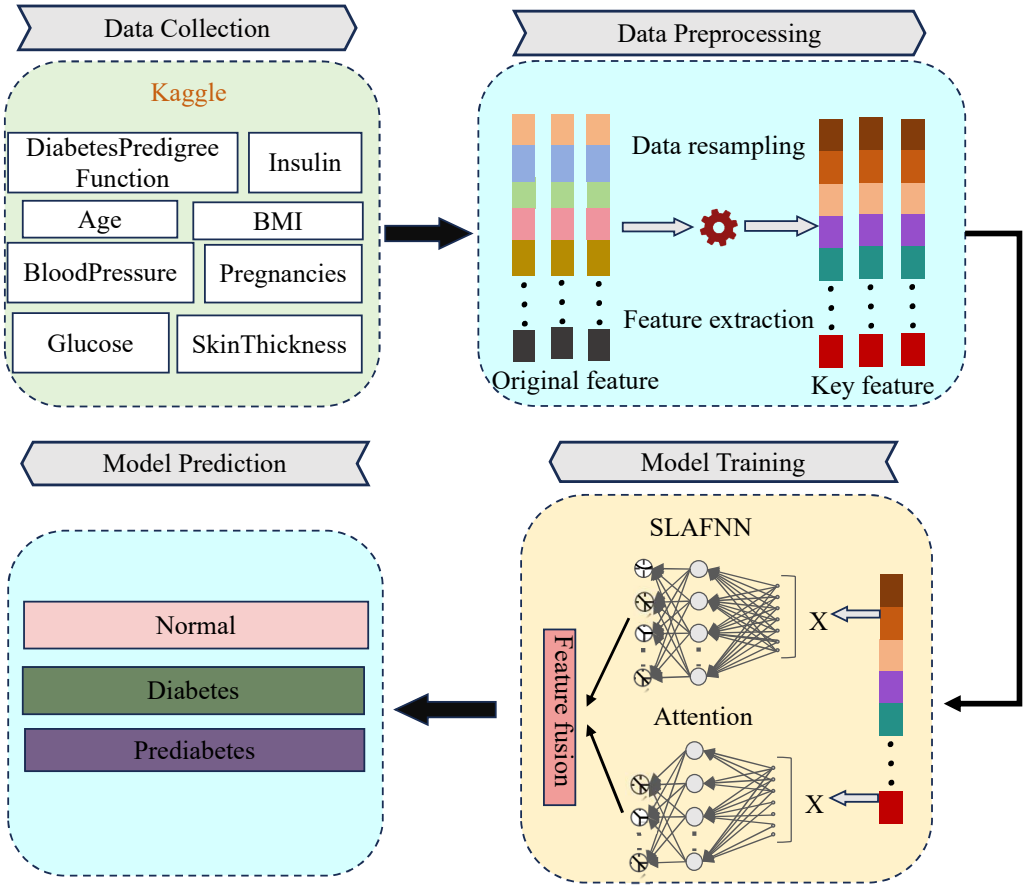

Figure 1. Diabetes prediction framework structure.

3.1. Self-learning activation neural network units

Activation functions are vital components in neural networks, playing a crucial role in deep learning. They transform thus input signals into output signals, thereby resembling the activation process in biological neurons. By introducing nonlinearity, activation functions enable neural networks to model nonlinear data. Without activation functions, neural networks would be limited to a series of linear transformations, thus only capable of representing linear functions. Additionally, activation functions have the ability to constrain output values within a specific range. For example, the sigmoid function limits its outputs between 0 and 1, while the tanh function restricts its outputs between −1 and 1.

This constraint makes network training more stable, thus preventing the problem of exploding or vanishing gradients. Certain activation functions (such as ReLU) can increase the sparsity of output signals, meaning that in most cases, the output value is zero and only a few non-zero outputs exist. This can reduce the computation costs and storage requirements and help address the issue of overfitting. The backpropagation algorithm in deep neural networks requires gradients to propagate through the network, and the size and direction of gradients are crucial for the network training. The choice of activation function can affect the speed and stability of the gradient flow, thus impacting the network training and performance. Traditional activation functions such as sigmoid and tanh functions tend to have very small gradients, even close to zero, when the input values are large or small. This makes it

difficult for gradients to propagate through deep networks, thus making them hard to train and possibly leading to the vanishing gradient problem.

An automatically learned activation function ACON [16] has been proposed, which can automatically learn whether to activate or what form of activation to use. This method of automatic learning greatly increases the flexibility of the activation function and model. The basic idea is to use the information learned by the neural network during training to dynamically adjust the shape of the activation function to adapt to different input distributions and task requirements. Suppose the current neural networks has an input $X = (x_1, x_2, x_3, \cdots, x_i)$ and an output $Y = (x_1, x_2, x_3, \cdots, x_j)$. The adaptive activation function $f(x)$ can be represented as the formula below:

$$f(x, \beta) = (p_1 - p_2)x \cdot \sigma[\beta(p_1 - p_2)x] + p_2 x \tag{3.1}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function, and β controls the smooth transition of the activation function. Motivated by this, we propose the SLALU for one-dimensional temporal medical data. The core computational formulas of SLALU are given as follows:

$$\begin{cases} g(x, \beta) = (p_1 - p_2)x \cdot \sigma(\beta \cdot (p_1 - p_2)x) + p_2 x \\ \beta(x) = \sigma(\text{mean}(x)), \quad \sigma(z) = \dfrac{1}{1 + e^{-z}} \end{cases} \tag{3.2}$$

where p_1 and p_2 are learnable parameters. Unlike the original ACON where β is a channel-wise global parameter, in SLALU, β is dynamically generated based on the mean value of the input features, thus enabling the activation shape to adaptively adjust according to the sample distribution. Figure 2(b) shows a visualization of different β values. When $\beta = 0$, the function behaves as a linear mapping and the neural network does not activate. $p_1 \& p_2$ control the degree of curvature of the function, as shown in Figure 2(a).

Given that existing activation functions are primarily applied in three-dimensional images and that the update of β is based on a channel-wise approach, the activation method for one-dimensional neural networks still largely relies on ReLU. Therefore, by optimizing β, this chapter aims to redefine the activation function. Updating the β parameter of a variable activation function through averaging offers several advantages:

1. **Improved Learning Efficiency**: Averaging can smooth out noise in the parameter updates, which may lead to more stable learning. It potentially allows β to converge to optimal values that generalize better across the dataset.
2. **Increased Stability**: By using the mean value, the updates to β become less sensitive to outliers and extreme values in the data, thus contributing to a more stable training process.
3. **Reduced Complexity**: Computational complexity is decreased as the update relies on a simple arithmetic mean rather than more complex operations, which can be particularly advantageous in large-scale neural networks.
4. **Resource Optimization**: Lower computational requirements also mean reduced memory usage and power consumption, which is valuable for the run-time efficiency and when deploying models to devices with limited resources.
5. **Faster Convergence**: The simplified computation through averaging may lead to a faster convergence of the training process, as it facilitates a smoother and more gradual updating mechanism.

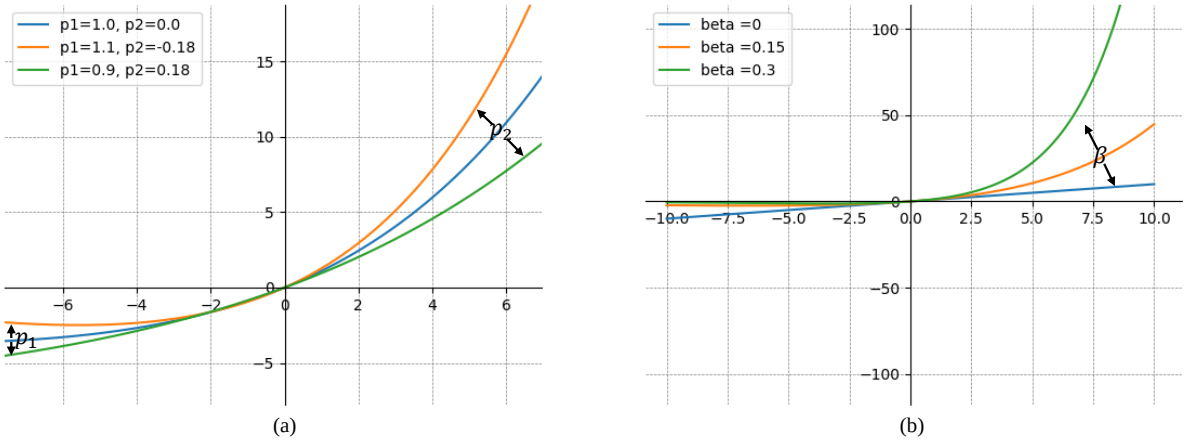

Figure 2. The visualization of learnable activation function.

Figure 3 shows how we applied this activation to the diabetes task. We observe the application of two mapping functions, f_1 and f_2, used to transform the input data to produce two different feature vectors Z_1 and Z_2. These mapping functions can be formalized as linear transformations $Z_i = W_i x + b_i$, where W_i and b_i represent the weight matrix and bias matrix of the linear transformation, respectively. By passing the input data x through these mappings, we obtain two sets of distinct feature representations.

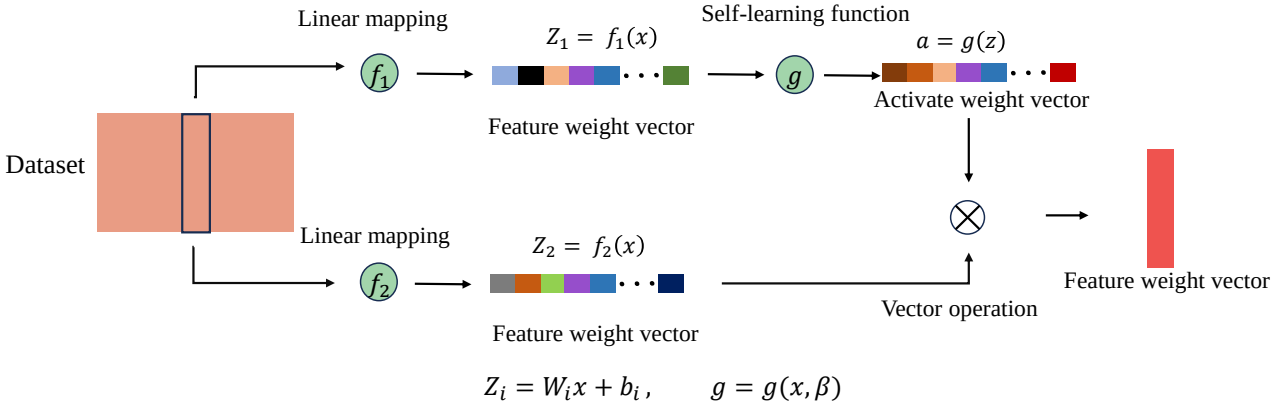

Figure 3. The visualization of learnable activation function computation process in one-dimensional diabetes dataset.

Subsequently, the feature vectors Z_1 and Z_2 are further transformed through an adaptive learning nonlinear function $g(x,\beta)$, to generate the final activation weight vector a. This adaptive function employs an optimized method $g(x,\beta)$, wherein β is a parameter adjusted during the learning process, which is calculated through a mean function.

Finally, this activation weight vector a is utilized in vector operations combined with Z_2, to produce the final output feature vector. This process is a common step in feature transformation $\sum_{i=1}^{n}(Z_i a + b_i)$ and extraction in deep learning, thus enabling the model to approximate complex function mappings by learning different levels of feature expressions.

3.2. Attention mechanism

To emphasize the correlation between different physiological indicators and to take their importance and mutual influence into account, we can introduce a self-attention mechanism into neural networks. The self-attention mechanism is a commonly used technique in deep learning that involves weighting the elements within an input sequence to better capture their inherent relationships.

By introducing a self-attention mechanism, the model can automatically learn the relationships and importance between elements, thus improving its performance and accuracy. The self-attention mechanism can help the model to better understand the internal structure of the input data, thus making it particularly useful in handling the correlation between physiological indicators. By performing a weighted summation, the self-attention mechanism allows the model to focus on other elements that are relevant to the current element, thus enhancing the model's expressive power and generalization ability.

If our input sequence is represented as $X = (x_1, x_2, x_3, \cdots, x_i)$, then the computation process of the self-attention mechanism can be described as follows. First, we transform the input sequence X into three matrices, namely Q, K, and V, through linear transformations where all three matrices have the same dimensions. Second, by calculating the similarities between the Q and K matrices, we obtain the attention matrix A with dimensions $d_{i\times i}$. The calculation process of A can be represented by the following formula:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)S \tag{3.3}$$

Next, we normalize the attention matrix A to obtain the weighted coefficient matrix W. Finally, we multiply the weighted coefficient matrix W with the V matrix to obtain the output matrix Z. The calculation process of matrix Z and matrix W is shown in the following formula:

$$W = \text{softmax}(A) \tag{3.4}$$

$$Z = W \cdot V \tag{3.5}$$

The $softmax$ function is indeed used to normalize each column in the attention matrix A. It ensures that the sum of each column is equal to 1. Finally, the attention matrix W is multiplied with the V matrix to obtain the output matrix Z. To validate the actual contribution of the attention module, we designed an ablation study: on the SLALU2 architecture, the attention layer was either removed or retained while keeping all other hyperparameters consistent, and the experiment was repeated five times. The results in Table 6 show that after removing the attention mechanism, the average accuracy drops to 91.82%, while the model that retains the attention layer achieves an average accuracy of 98.18%. Furthermore, the AUC decreased from 0.986 to 0.941, thus confirming the necessity of the attention mechanism for feature reweighting. In addition, Figure 5 no longer displays a generic attention structure; it was replaced by a heatmap of the attention weights learned after training (see the new Figure 5), which intuitively shows the physiological indicators that the model focuses on.

4. Experiment

This session provides a detailed introduction to our experimental parameter settings and two key experiments. We chose an MLP neural network as the main structure of our network, combined with a

layer of an attention mechanism as the experimental topology, with the specific structure as shown in Table 2. These experiments were designed to evaluate the impact of SLALU versus traditional activation functions (ReLU, PReLU, and Swish) on the accuracy of diabetes prediction across different network structures and training iterations. Before the experiments, we conducted thorough data preprocessing. We opted for a simple feedforward neural network as our research framework.

Table 2. The Context-Adaptive activation network topology.

Dense and attention layer	(Input, Output)	Number
Attention	(4, 4)	1
SLALU	(4, 4)	1
Dense	(8, 8)	1
SLALU	(8, 8)	1
Dense	(16, 16)	1
SLALU	(16, 16)	1
Dense	(3, 3)	1
Softmax	(3, 3)	1

For networks that employed SLALU units, we chose to randomly initialize $P_1\&P_2$ parameters randomly, setting the parameters to the mean values of the features. Specific experimental parameter settings are shown in Table 3.

Table 3. The table is the parameter settings for diabetes prediction.

Hyperparameter	Number
Epoch	200
Learning Rate	0.005
Optimizer	Rmsprop
Number of Dense	3
Number of Attention layers	1
Batch size	32
Loss	Cross-Entropy-L2
Layer Activation	SLALU, ReLU, PReLU, Swish

In both sets of experiments, we used the TensorFlow library for all testing. In the first experiment, we explored the effect of different activation functions on the accuracy of diabetes prediction using the same network structure. The detailed display of the number of hidden nodes in each layer of the network structure is shown in Table 4, where H_n represents the number of hidden nodes in the layer indexed by n, and R^n represents the number of hidden nodes.

Additionally, we introduced an attention mechanism to enhance the network's sensitivity to different features and effectively integrate these features. The second experiment aimed to compare the impact of different activation functions on the network convergence efficiency after 80 training iterations. These experiments all tested the role of activation functions in the network training and their contribution to the convergence speed. All related experimental codes can be found at the following link: https://github.com/Hike688/SLALU.git.

Table 4. Different architectures.

Architecture	Number of neurons in each hidden layer
SLALU1	$H_1 \in R^{16}$
SLALU2	$H_1 \in R^8$, $H_2 \in R^{16}$
SLALU3	$H_1 \in R^8$, $H_2 \in R^{16}$, $H_3 \in R^8$
SLALU4	$H_1 \in R^{16}$, $H_2 \in R^{16}$, $H_3 \in R^8$
SLALU5	$H_1 \in R^8$, $H_2 \in R^8$, $H_3 \in R^{16}$, $H_4 \in R^{16}$
SLALU6	$H_1 \in R^8$, $H_2 \in R^8$, $H_3 \in R^{16}$, $H_4 \in R^8$
SLALU7	$H_1 \in R^8$, $H_2 \in R^{16}$, $H_3 \in R^{16}$, $H_4 \in R^8$

4.1. Data preprocessing

4.1.1. Missing value and outlier handling

Missing value and outlier handling involves two steps: quartile-based outlier detection and outlier imputation. Quartile-based outlier detection is a commonly used method to handle outliers and is widely applied in data preprocessing.

The basic idea is to determine whether a data point is an outlier by calculating the quartiles (Q_1, Q_2, Q_3) and the interquartile range ($IQR = Q_3 - Q_1$) of the data. Typically, data points that exceed the upper limit ($Q_3 + 1.5IQR$) or the lower limit ($Q_1 - 1.5IQR$) are considered outliers and are processed accordingly. In practice, tools such as pandas and numpy in Python can be used for quartile-based outlier detection. By reading the raw data file, we can calculate the Q_1, Q_2, Q_3, and IQR statistics for each column of data.

Based on the aforementioned formulas, we can determine the upper and lower limits for outlier detection and mark the data points as True or False to indicate whether they are outliers. In the experimental section of this paper, as shown in Figure 4, mean imputation is employed as a common method to handle outliers, where the outliers are replaced with the mean of the corresponding column. This approach helps alleviate the impact of outliers on the results of model training; after identifying all outliers through quartile-based outlier detection, the mean of non-outlier data points in each column is calculated and used to replace the outliers.

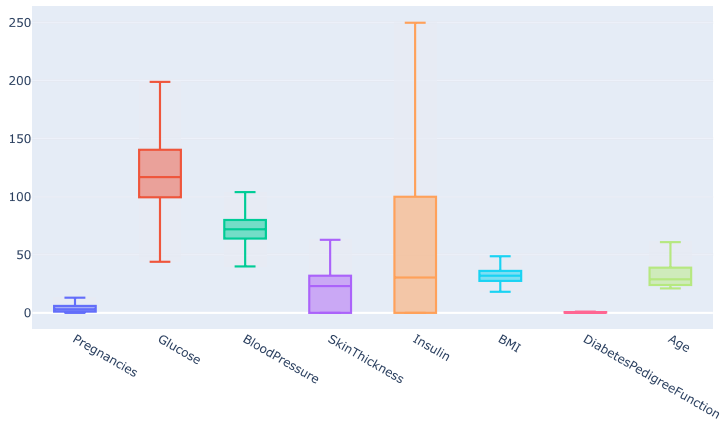

Figure 4. The figure represents the distribution of outliers in the Pima dataset after applying quartile-based outlier detection and replacing them with the mean.

4.1.2. Feature selection for diabetes prediction

A Spearman correlation is a commonly used statistical method to measure the correlation between two variables. It can not only detect linear relationships but also assess nonlinear associations, thus making it a very useful tool in various scenarios. Compared to other methods, a Spearman correlation is a non-parametric method and is not constrained by the distribution of the data.

In a Spearman correlation, a correlation coefficient closer to 1 indicates a strong positive correlation between two variables. This means that when one variable increases, the other variable usually increases as well, and their ranks correspondingly increase. If the correlation coefficient falls between 0.5 and 1, it signifies a weak positive correlation, where an increase in one variable tends to be associated with an increase in the other variable, though the correlation is not strong. When the correlation coefficient approaches 0, it indicates no linear relationship between the two variables, and there is no clear association between their ranks. If the correlation coefficient ranges between -1 and -0.5, it suggests a weak negative correlation, where an increase in one variable tends to be associated with a decrease in the other variable. Finally, a correlation coefficient close to -1 indicates a strong negative correlation, where an increase in one variable usually corresponds to a decrease in the other variable, and their ranks decrease accordingly.

For features in the diabetes dataset (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, and Age), we conducted a Spearman correlation analysis using the SPSS tool [34]. Based on the results reported from the Spearman analysis, we can conclude that Glucose, BloodPressure, Insulin, and Age are representative features in the diabetes dataset. This suggests that these features may play important roles in studying the correlation with diabetes and are worthy of further investigation and analysis.

4.2. Accuracy of different activation methods

In our first experiment, we focused on evaluating the performance of the SLALU in neural network structures. This network first passes through one layer of attention mechanisms, followed by MLP. We specifically focused on the characteristic that the SLALU's parameter β is dynamically updated based on the mean of the input features to adapt to different data distributions.

By comparing with traditional activation functions (such as ReLU, PReLU, and Swish), our aim was to explore the impact of these different activation functions on the network performance, with a particular focus on their performance in simplified network environments without complex structures like the residual connections. To ensure the consistency of the experiments, all models were initialized using the same method and trained for 200 epochs using a stochastic gradient descent with the RMSprop optimizer.

As shown in Table 5, SLALU demonstrated the best or near-best performance across all tested network architectures. SLALU reached an accuracy of 98.18% in both the SLALU2 and SLALU6 architectures, which is significantly higher than the results using ReLU and PReLU in the same architectures. This highlights the importance of SLALU's dynamic adjustment of activation thresholds to adapt to input data features in enhancing the network performance. Especially in simpler scenarios, SLALU exhibited a good accuracy. According to our data, SLALU achieved better accuracies in most test architectures, particularly in SLALU2, SLALU3, SLALU4, SLALU5, and SLALU6, thereby showing accuracies of 98.18%, 96.36%, 96.36%, 97.27%, and 98.18%,

respectively, thus indicating its suitability for complex machine learning tasks such as diabetes prediction. SLALU performed better or at least comparably in all architectures relative to traditional ReLU and its variant PReLU, especially when ReLU and PReLU performed poorly, such as ReLU's mere 91.82% accuracy in SLALU5.

Table 5. Comparison of different activation function approaches on accuracy.

Architecture	SLALU	ReLU	PReLU	Swish
SLALU1	94.55%	**96.36%**	94.55%	94.55%
SLALU2	**98.18%**	95.45%	93.64%	94.55%
SLALU3	**96.36%**	95.45%	92.72%	94.55%
SLALU4	**96.36%**	95.45%	94.55%	95.45%
SLALU5	**97.27%**	91.82%	93.64%	93.64%
SLALU6	**98.18%**	**92.73%**	90.91%	95.45%
SLALU7	94.55%	90.91%	95.45%	**96.36%**

Although the Swish activation function exhibited good performances in some architectures, it generally did not surpass SLALU, particularly in SLALU5, where Swish dropped to a 93.64% accuracy.

While PReLU showed improvements over ReLU in some architectures, reflecting the potential value of activation functions that are unsaturated on both sides, SLALU went even further by making significant contributions to the model accuracy through its adaptive parameter update mechanism. Although the Swish activation function performed well in certain architectures, overall, it still did not match the comprehensive performance of SLALU. Moreover, SLALU's outstanding performance demonstrated the effectiveness of introducing parametric features and adaptive mechanisms into the design of activation functions. By updating based on the mean of features, the SLALU activation function not only has a strong adaptability but also significantly enhances the performance for complex tasks such as processing through multiple layers of a feedforward network, which was fully validated and showcased in our experimental design.

4.3. Ablation Study and Attention Weight Visualization

To further validate the contribution of the attention mechanism, we conducted an ablation study that compared SLALU2 without attention (baseline) and SLALU2 with the proposed Atten_model. The experimental results are summarized in Table 6. As shown, the attention-enhanced model achieves a significantly higher accuracy (98.18% vs. 91.82%) and AUC (0.986 vs. 0.941), thus confirming the necessity of adaptive feature reweighting.

Table 6. Ablation study: SLALU2 with and without attention mechanism.

Model	Acc (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	AUC
SLALU2	91.82	92.07	96.06	90.70	91.15	0.941
SLALU2 (with attention)	98.18	96.81	98.45	97.77	97.25	0.986

The substantial performance gap indicates that the attention layer enables the network to focus on discriminative features. Without attention, the fully connected layers treat all input features equally, thus leading to suboptimal representation learning. In contrast, the attention module adaptively

amplifies critical physiological indicators (e.g., Glucose) while suppressing irrelevant ones, thus boosting both the accuracy and the AUC.

Attention weight visualization: To gain insight into the attention mechanism, we visualized the learned weight matrices as heatmaps (input features: Glucose, BloodPressure, Insulin, Age), as shown in Figure 5.

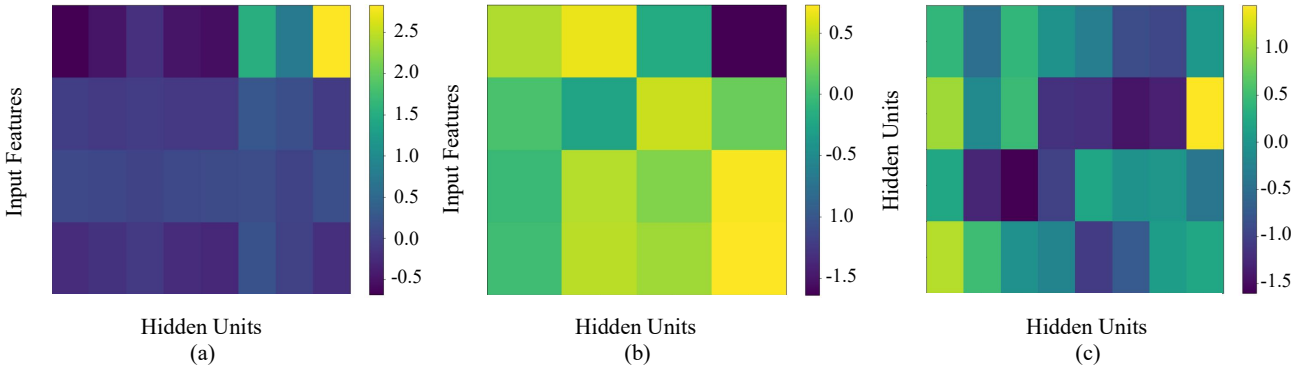

Figure 5. Attention weight visualization.

Three representative heatmaps are analyzed:

1. **SLALU2 (no attention, Figure 5(a)):** The weights are flat with low discrimination. Only the first feature (Glucose) shows slightly higher weights in a few channels; the other three features remain near-zero (dark blue). This confirms that a vanilla fully connected layer cannot autonomously identify feature importance.
2. **SLALU2 (attention layer, Figure 5(b)):** The weight distribution becomes polarized. For Glucose (Y=0 row), some channels receive high positive weights (bright yellow) while others are suppressed to negative values (dark purple). Moreover, the other three features exhibit clear differentiation. Hence, the attention layer selectively enhances task-relevant features and suppresses noise.
3. **SLALU2 (post-attention fully connected layer, Figure 5(c)):** Compared to SLALU2(no attention), the weight distribution is structured and hierarchical. High-weight regions (bright yellow/green) and low-weight regions (dark purple) clearly emerge. For instance, the second feature (BloodPressure) shows a notable positive peak at channel 7. This demonstrates that the attention layer provides a useful feature-importance prior, thus enabling subsequent layers to learn more efficiently and to better suppress irrelevant features (e.g., Insulin, Age).

Key role of the attention layer: From the above analysis, the attention layer plays three critical roles in diabetes classification: (i) *active feature focusing* – it adaptively assigns higher weights to key features such as Glucose; (ii) *improved feature utilization* – it provides a prior for subsequent fully connected layers, thus leading to more structured and efficient learning; and (iii) *enhanced interpretability* – the heatmap visualization clearly reveals the model's decision basis, thus increasing clinical trust.

4.4. Convergence rate of different activation methods

In our second experiment, we delved deeper into the efficacy of various activation functions within the same neural network structure as shown in Table 2 and on the same task, aiming to highlight the advantages of our SLALU activation function. We compared three commonly used activation functions against our novel SLALU activation function, thereby applying data preprocessing and a Spearman-based feature selection before initiating the network training. To ensure consistency and isolate the impact of the activation functions from other variables, we maintained identical conditions across all other parameters. The networks were trained for 80 iterations using a stochastic gradient descent with the RMSprop optimizer. This approach allowed us to closely observe and compare the initial convergence speeds of the different activation functions, with the learning curves plotted for each across the testing dataset during training illustrated in Figure 6.

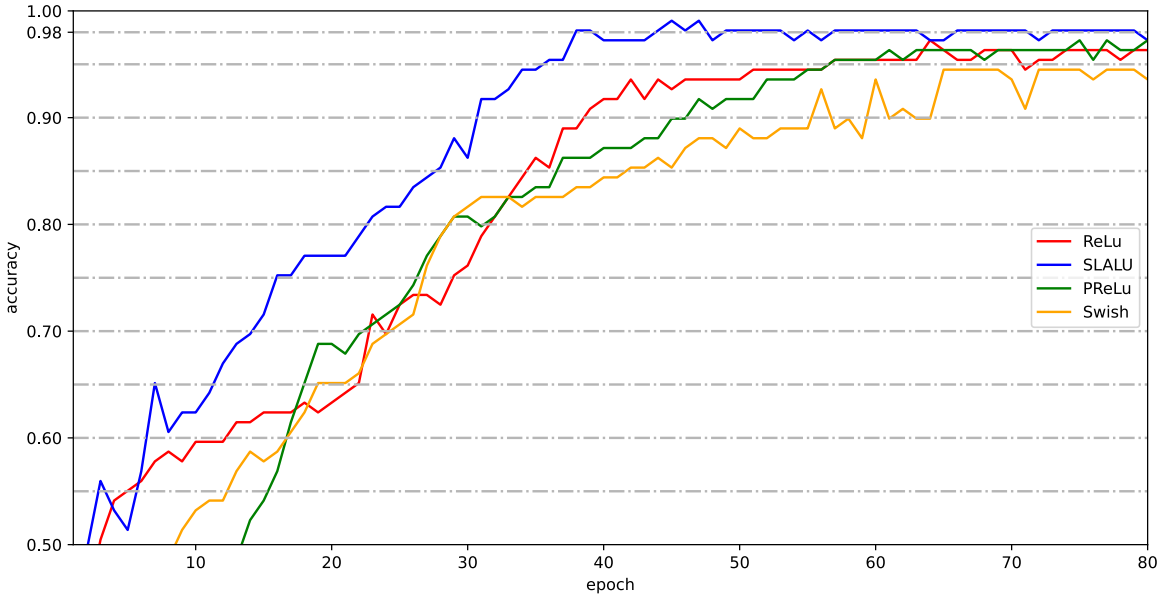

Figure 6. Comparison of convergence speeds of different activation function approaches on the validation set.

Figure 6 clearly presents the performance of the four activation functions under identical neural network settings and tasks. It is evident that the convergence curve for our proposed SLALU activation function (depicted as the blue line) achieves a notable convergence by the 50th iteration. Not only does it converge faster than the alternatives, but it also exhibits a higher accuracy rate. This difference in performance indicates the advantage of our SLALU activation function, particularly in complex machine learning tasks such as diabetes prediction.

By analyzing the convergence patterns depicted in Figure 6, we gain deep insights into the dynamic and adaptive nature of the SLALU activation function. Unlike traditional activation functions that offer static responses across all iterations of training, the SLALU function adaptively adjusts its parameters based on the mean of the input features. This self-adjusting mechanism ensures that SLALU maintains an optimal balance between allowing necessary information flow and mitigating the vanishing gradient problem, which is a common pitfall for many activation functions. In the context of diabetes prediction, where the dataset's features may exhibit significant variability and complexity, SLALU's ability to

dynamically adjust becomes critically important. Furthermore, the quicker convergence and higher accuracy of SLALU compared to its counterparts can be attributed to its enhanced capability to navigate through complex, high-dimensional data landscapes. This is particularly vital in early training phases where efficient navigation through the parameter space can substantially accelerate the learning process and lead to a higher overall model performance.

Hence, the experiment's results provide compelling evidence of the SLALU activation function's value, not just in speeding up the convergence process but in achieving superior accuracies in predictive tasks. This demonstrates that our SLALU function represents a significant advancement in the design of neural network activation functions, particularly for applications in healthcare analytics where the predictive accuracy is of paramount importance.

On the PIDD, our method achieved an accuracy of 98.18%, thus representing a significant improvement compared to currently outstanding diabetes models in terms of the accuracy, as shown in Table 7.

Table 7. Comparison of diabetes prediction accuracy among different methods.

Author	Method	Accuracy
Ours	Self-learning Action Linear Unit	98.18%
Zhuang et al. [21]	CNN-Voting ensemble (CNN + soft voting)	98%
Whig et al. [23]	Pycaret machine learning technology	90%
Smith et al. [25]	ADAP neural network	76%
Saxena et al. [26]	Random Forest	79.8%
Bouhissi et al. [27]	Voting ensemble (LR, SVM, KNN, GBC, NB, RF)	82.0%
Chou et al. [28]	Two-Class Boosted Decision Tree	95.3%
Waughfa et al. [30]	Multilayer Perceptron (MLP)	85%
Soltanizadeh et al. [31]	Hybrid CNN+MLP	93.89%
Ateequr Rahaman et al. [32]	Ensemble (Random Forest, Boosting, Bagging, Stacking)	88%

5. Conclusion

In this work, we proposed a SLALU-based deep learning framework to improve the performance of diabetes prediction. The framework integrates data preprocessing methods, an adaptive activation linear unit, neural networks, and attention mechanisms. The framework is applicable to this proposed research model, which delivers excellent results in assessing diabetes. In this study, we demonstrated through the first experiment that our proposed SLALU, when applied to neural networks, enhances the network's fitting ability compared to popular activation functions. In the second experiment, contrasting with traditional activation functions, we showed that our SLALU, outperforms popular activation functions in terms of the convergence speed. Therefore, the method proposed in this paper provides guidance for the future prediction of diabetes. Nevertheless, this study has the following limitations: (1) it relies on a single age-region population (Pima), and the external validation performance shows a certain degree of decay; (2) SLALU introduces additional parameters p_1, p_2 and a dynamic β computation, which increases the forward propagation time compared to ReLU; (3) the clinical interpretability remains relatively weak, as although attention weights can demonstrate feature

importance, they cannot provide causal explanations. In future work, we will explore lightweight adaptive activation functions and combine SHAP values to provide individualized prediction bases.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

This research is partially supported by National Natural Science Foundation of China (61972438) and Wuhu Science and Technology Plan Project (2023yf117). The authors would like to thank their colleagues and students at Anhui Provincial Engineering Research Center of Medical Big Data Intelligent System.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

Ren Chen proposed the conceptualization and reviewed & edited the manuscript. Chao Wang conducted the experiment and wrote the original draft. Fulong supervised this study and validated it. All authors reviewed the manuscript.

Novelty claim

The proposed novel diabetes prediction framework underpinned by a neural network architecture featuring a parameterized activation function, called the Self-Learning Activation Linear Unit (SLALU), is a classifier model equipped with an automatically adjustable activation neural network alongside a one-dimensional self-attention mechanism.

Ethics approval of research

The Ethics Committee of Anhui Normal University, approved all procedures. The datasets PIDD and LMCH for this study can be found in Kaggle (www.kaggle.com).

References

1. Khan R, Chua Z, Tan J, et al. (2019) From pre-diabetes to diabetes: diagnosis, treatments and translational research. *Medicina* 55: 546. https://doi.org/10.3390/medicina55090546
2. Vaishali R, Sasikala R, Ramasubbareddy S, et al. (2017) Genetic algorithm based feature selection and moe fuzzy classification algorithm on pima indians diabetes dataset,

International Conference on Computing Networking and Informatics (ICCNI), 1–5. https://doi.org/10.1109/ICCNI.2017.8123815

3. Cheng Y and Caughey A (2008) Gestational diabetes: diagnosis and management. *J Perinatol* 28: 657–664. https://doi.org/10.1038/jp.2008.62
4. Bukhari M, Alkhamees B, Hussain S, et al. (2021) An improved artificial neural network model for effective diabetes prediction. *Complexity* 2021: 5525271. https://doi.org/10.1155/2021/5525271
5. Khanam J and Foo S (2021) A comparison of machine learning algorithms for diabetes prediction. *Ict Express* 7: 432–439. https://doi.org/10.1016/j.icte.2021.02.004
6. Naz H and Ahuja S (2020) Deep learning approach for diabetes prediction using pima indian dataset. *J Diabetes Metab Disord* 19: 391–403. https://doi.org/10.1007/s40200-020-00520-5
7. Alam T, Iqbal M, Ali Y (2019) A model for early prediction of diabetes. *Inf Med Unlocked* 16: 100204. https://doi.org/10.1016/j.imu.2019.100204
8. Fröhlich W, Rigo S, Bez M (2024) Athena i—an architecture for near real-time physiological signal monitoring and pattern detection. *Future Gener Comp Sy* 150: 395–411. https://doi.org/10.1016/j.future.2023.09.010
9. Liu J, Xiao Z, Lu S, et al. (2023) Infrastructure-level support for gpu-enabled deep learning in dataview. *Future Gener Comp Sy* 141: 723–737. https://doi.org/10.1016/j.future.2022.12.014
10. Sivaranjani S, Ananya S, Aravinth J, et al. (2021) Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction, *7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1: 141–146. https://doi.org/10.1109/ICACCS51430.2021.9441935
11. Gnanadass I (2020) Prediction of gestational diabetes by machine learning algorithms. *IEEE Potentials* 39: 32–37. https://doi.org/10.1109/MPOT.2020.3015190
12. Hasan M, Alam M, Das D, et al. (2020) Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access* 8: 76516–76531. https://doi.org/10.1109/ACCESS.2020.2989857
13. Dwivedi AK (2018) Analysis of computational intelligence techniques for diabetes mellitus prediction. *Neural Comput Appl* 30: 3837–3845. https://doi.org/10.1007/s00521-017-2969-9
14. Shen L, Chen H, Yu Z, et al. (2016) Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl-Based Syst* 96: 61–75. https://doi.org/10.1016/j.knosys.2016.01.002
15. Wu Y , Zhang Q , Hu Y, et al. (2022) Novel binary logistic regression model based on feature transformation of XGBoost for type 2 diabetes mellitus prediction in healthcare systems. *Future Gener Comp Sy* 129: 1–12. https://doi.org/10.1016/j.future.2021.11.003
16. Ma N, Zhang X, Liu M, et al. (2021) Activate or not: learning customized activation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8032–8042. https://doi.org/10.1109/CVPR46437.2021.00794
17. Nair V and Hinton G (2010) Rectified linear units improve restricted boltzmann machines, *27th international conference on machine learning (ICML)*, 807–814. https://doi.org/10.5555/3104322.3104425

18. He K, Zhang X, Ren S, et al. (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification, *Proceedings of the IEEE international conference on computer vision*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123
19. Godfrey L and Gashler M (2015) A continuum among logarithmic, linear, and exponential functions, and its potential to improve generalization in neural networks, *7th international joint conference on knowledge discovery, knowledge engineering and knowledge management (IC3K)*, 1: 481–486. https://doi.org/10.48550/arXiv.1602.01321
20. Trottier L, Giguere P, Chaib-Draa B (2017) Parametric exponential linear unit for deep convolutional neural networks, *16th IEEE international conference on machine learning and applications (ICMLA)*, 207–214. https://doi.org/10.1109/ICMLA.2017.00038
21. Zhuang K, Zhang C, Chen Z, et al. (2025) Integrating convolutional neural networks with ensemble methods for enhanced diabetes diagnosis: a multi-dataset evaluation. *Front Med* 12: 1657889. https://doi.org/10.3389/fmed.2025.1657889
22. Tian G, Yang Y, Wen S (2025) Time-series stock price forecasting based on neural networks: a comprehensive survey. *Artif Intell Sci Eng* 1: 255–277. https://doi.org/10.23919/AISE.2025.000018
23. Whig P, Gupta K, Jiwani N, et al. (2023) A novel method for diabetes classification and prediction with pycaret. *Microsyst Technol* 29: 1479–1487. https://doi.org/10.1007/s00542-023-05473-2
24. AlZu'bi S, Elbes M, Mughaid A, et al. (2023) Diabetes monitoring system in smart health cities based on big data intelligence. *Future Int* 15: 85. https://doi.org/10.3390/fi15020085
25. Smith JW, Everhart JE, Dickson WC, et al. (1988) Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, *Proceedings of the annual symposium on computer application in medical care*, 9: 261–265.
26. Saxena R, Sharma SK, Gupta M, et al. (2022) A novel approach for feature selection and classification of diabetes mellitus: machine learning methods. *Comput Intel Neurosce* 2022: 3820360. https://doi.org/10.1155/2022/3820360
27. Mushtaq Z, Ramzan MF, Ali S, et al. (2022) Voting classification-based diabetes mellitus prediction using hypertuned machine-learning techniques. *Mob Inf Syst* 2022: 6521532. https://doi.org/10.1155/2022/6521532
28. Chou CY, Hsu DY, Chou CH (2023) Predicting the onset of diabetes with machine learning methods. *J Pers Med* 13: 406. https://doi.org/10.3390/jpm13030406
29. Al Reshan MS, Amin S, Zeb MA, et al. (2024) An innovative ensemble deep learning clinical decision support system for diabetes prediction. *IEEE Access* 12: 106193–106210. https://doi.org/10.1109/ACCESS.2024.3436641
30. Waughfa ZM, Adnan II, Sultana SS, et al. (2024) Diabetes prediction: a comprehensive study integrating deep learning and machine learning approaches, *Proceedings of the 3rd International Conference on Computing Advancements*, 520–526. https://doi.org/10.1145/3723178.3723247
31. Soltanizadeh S, Mobini M, Naghibi SS (2025) Design of a low-complexity deep learning model for diagnosis of type 2 diabetes. *Curr Diabetes Rev* 21: E15733998307556. https://doi.org/10.2174/0115733998307556240819093038

32. Rahaman MA, Idris RM, Zuveriya S, et al. (2024) Enhancing diabetes mellitus onset prediction through advanced ensemble learning techniques. *JOSMA* 6: 11–28. https://doi.org/10.22452/josma.vol6no2.2

33. Andrade-Arenas L and Yactayo-Arias C (2025) Comparative evaluation of machine learning models for diabetes prediction: a focus on ensemble methods. *Ingenierie des Systemes d'Information* 30: 1795. https://doi.org/10.18280/isi.300712

34. Landau S and Everitt B (2003) *A Handbook of Statistical Analyses using SPSS*, New York: CRC Press. https://doi.org/10.1201/9780203009765

AIMS Press