



## Diabetes Prediction Using Parametric Swish-based Recurrent Neural Network

Suhas Kamshetty Chinnababu<sup>1\*</sup>      Anandababu Jayachandra<sup>1</sup>

<sup>1</sup>*Department of Information Science and Engineering, Malnad College of Engineering,  
Hassan and Visvesvaraya Technological University, Belagavi, India*

\* Corresponding author's Email: [suhas2385@gmail.com](mailto:suhas2385@gmail.com)

---

**Abstract:** Diabetes is a globally dangerous disease characterized by decreased insulin levels and increased blood sugar which causes significant health risks and requires careful management. However, the class imbalance issue leads to a biased model resulting in poor prediction performance due to insufficient pattern learning. In this research, the Parametric Swish-based Recurrent Neural Network (PSRNN) is proposed to accurately and effectively predict diabetes. Initially, the data is gathered from the Pima Indian Diabetes Dataset (PIDD) and Frankfurt Hospital, Germany datasets to analyze the model's performance. Then, min-max normalization is used to normalize the data by scaling process which preserves the relative relationships among data points. Then, the Adaptive Synthetic Technique (ADASYN) is employed to balance the imbalanced data in diabetes prediction which improves the model's performance by generating synthetic minority class samples. At last, PSRNN is performed for diabetes prediction. The PSRNN achieves a better accuracy of 99.81% and 0.992 on the PIDD and Germany datasets, in comparison to the existing techniques like Support Vector Machine (SVM), Deep Convolutional Neural Networks (DCNN), and Artificial Neural Networks (ANN).

**Keywords:** Adaptive synthetic technique, Deep convolutional neural networks, Hard swish rectified linear unit-recurrent neural network, Min-max normalization, Pima Indian diabetes dataset.

---

### 1. Introduction

Diabetes is a chronic disease that directly impacts the pancreas by inhibiting the production of insulin. Insulin is primarily responsible for managing the glucose levels. Numerous factors like physical inactivity, excessive body weight, abnormal cholesterol levels, and high blood pressure cause a person to be easily affected by diabetes [1]. Diabetes contains four clinical types that are, type I, type II, prediabetes, and gestational diabetes. Type I diabetes transpires within the body while the immune system damages pancreatic beta cells which produce only a tiny amount of insulin or no insulin at all [2, 3]. Type II is a metabolic disorder in which the body's cells become resistant to insulin, or when the pancreas does not produce enough insulin to regulate the blood glucose levels [4, 5]. Increased sugar in the blood is associated with death risk in the community because of acute myocardial infarction, stroke, pneumonia,

and so on. During pregnancy, gestational diabetes occurs and its symptoms subside as the pregnancy ends [6]. In the prediabetes stage, the blood glucose level stands above normal [7] with harmful effects on the primary organs, making it a significant contributing factor to many other diseases [8].

Diabetes is a metabolic condition that arises due to various health complications that cause millions of deaths every year all over the world. By 2030, the number of people with diabetes in evolving countries is expected to increase from 84 to 228 million, establishing a significant load on all healthcare system across the world [9]. Identifying essential features is crucial in controlling this disease for providing an effective treatment for affected individuals [10, 11]. The utilization of Deep Learning (DL) methods is driven by an abundance of available diabetes data which assists clinicians in identifying patients at high risk of developing diabetes complications to make informed treatment decisions. This represents a significant aid in solving the clinical

inertia and enhancing the quality of diabetes care [12, 13]. DL is a subset of Machine Learning (ML) which plays a major role in the time-series prediction field [14, 15]. However, the class imbalance issue leads to a biased model which results in poor prediction performance due to inadequate pattern learning that minimizes the accurate prediction performance. To overcome this issue, the PSRNN is proposed to accurately and effectively predict diabetes with the help of solving the imbalance issue using ADASYN. It adaptively adjusts the sampling depending on the local data density, creating a more representative training set that enhances the predictive model accuracy for diabetes prediction.

The main contributions of this research are as follows:

- Min-max normalization scales data to preserve the relative relationships among the data points which results in effective diabetes prediction.
- ADASYN balances the imbalanced data by generating synthetic samples for a minority class which enhances the accuracy of diabetes prediction.
- PSRNN is established to predict diabetes accurately and effectively. The RNN captures temporal dependencies' inpatient data, generating context for more accurate predictions. The Parametric swish activation allows the model to learn the sigmoid coefficient's optimal value, reducing vanishing gradients in neural networks, hence leading to faster training. The PSRNN effectively predicts diabetes by efficaciously executing this process.

The rest of the paper is organized as follows: Section 2 demonstrates the literature survey, Section 3 explains the proposed methodology in detail, while Section 4 indicates the results for the existing methods and the proposed method. Finally, Section 5 summarizes the overall conclusion of the paper.

## 2. Literature survey

The related work on diabetes prediction are discussed along with their major advantages and disadvantages.

Arora [16] suggested a K-means clustering with a Support Vector Machine (SVM) to predict diabetes patients. The K-means technique was used to pre-process the data and then, the diabetes prediction was carried out using SVM. The suggested approach enhanced the system's accuracy in the early diagnosis of diabetes by utilizing the clustering capabilities of the K-means to preprocess data. However, K-means

clustering was performed depending on the Euclidean distance by considering the isotropic clusters which made it less effective in capturing complex distribution of data due to variations in the clusters' size and shape.

Patro [17] implemented a Deep Convolutional Neural Network (DCNN) for accurate diabetes prediction. Initially, the pre-processing stage was employed to remove duplicate instances, missing values, outliers, and inconsistencies for a better understanding. Further, data modeling was performed based on the correlation measures among features, utilized to generate an effective data. Finally, DCNN was employed to predict the diabetes accurately. The data modeling approach significantly enhanced the model's performance by effectively capturing complex relationships within data. However, the class imbalance issue led to a biased model due to inadequate pattern learning which minimized the prediction performance.

Azbeq [18] developed a statistical predictive approach to predict diabetes effectively by utilizing the Pima Indian Diabetes Dataset (PIDD) and Hospital Frankfurt Germany datasets. Then, the data was scaled and filtered by employing noise-invariant data expansion. Finally, an adaptive Random Forest (RF) was used for data training and diabetes prediction. The developed approach performed well because of data management, analysis, and storage processes. However, the statistical approach struggled with capturing complex interactions between variables as it was based on linear relationships which led to inaccurate prediction.

Gourisaria [19] introduced an Artificial Neural Network (ANN) to predict diabetes mellitus. The data was obtained from the University of California, Irvine repository and Germany datasets. The dimensionality reduction approaches like Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used to enhance the model's performance, while minimizing the model's computational expense. However, ANN failed to recognize the patterns that were invariant to spatial transformations as they lacked explicit mechanisms to capture spatial relationships within the data.

Edeh [20] presented RF to provide an accurate treatment for diabetes patients in the early stage. The K-means approach was used for data correction to enhance the model's performance and control the classified technique. Then, the presented RF technique was performed for prediction. The quality of data was improved using K-means algorithm which enhanced the outcomes.

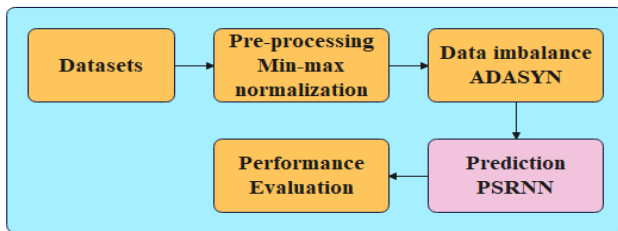


Figure. 1 Block diagram for the proposed technique

However, the RF struggled with interpretability due to the complexity of multiple tree decisions.

In the overall analysis, the existing techniques had limitations like class imbalance issues leading to a biased model due to inadequate pattern learning, and struggled to capture non-linear relationships and complex patterns. To overcome this issue, the PSRNN is proposed to effectively predict diabetes by solving class imbalance issues using the ADASYN technique.

### 3. Proposed methodology

The PSRNN is proposed for diabetes prediction in this research. Initially, data is obtained from two benchmark datasets: PIDD and Frankfurt Hospital, Germany datasets to determine the performance of PSRNN. The min-max normalization is employed to pre-process the obtained data by scaling operation which improves the interpretation of the model. Then, the ADASYN is performed to balance the imbalance data and finally, PSRNN is developed for diabetes prediction. Fig. 1 illustrates a block diagram for the PSRNN technique.

#### 3.1 Datasets

In this research, two benchmark datasets namely, PIDD dataset [21] which is provided by the UCI Machine Learning Repository and Frankfurt Hospital, Germany [22] are used to evaluate the proposed technique's performance. PIDD has data from 768 female patients with 8 features and 1 output with the patient's labels of pregnancies, blood pressure, glucose, insulin, Body Mass Index (BMI), skin thickness, age, diabetes pedigree function, and outcome. Likewise, PIDD dataset, and the Germany dataset contain the same 8 features and 1 output with the patient's label. It has 67.6% of the records being non-diabetes, while the remaining 32.4% are diabetes. Here, it is taken into consideration that every patient is female aged between 21 and 81. The gathered data are fed into the pre-processing stage for data scaling.

#### 3.2 Pre-processing

After obtaining data, the min-max normalization [23] is used to normalize the data by scaling process which preserves the relative relationships among data points for diabetes prediction. Normalizing the data reduces the model's complexity for further processing. It enables significant benefits for predicting the model when related to neural networks. The normalization of input values increases the training stage if the back-propagation approach is employed in a neural network which provides an effective neural network. The normalization function depends on data scaling that contains a min-max technique that is capable of converting the present data to fit within range of  $[-1, 1]$  and  $[0, 1]$  intervals. The formula for min-max normalization is provided in Eq. (1).

$$p = \frac{(x-x_{min})(max-min)}{(x_{max}-x_{min})+min} \quad (1)$$

Where,  $min$  and  $max$  indicate the specified input variable range,  $x_{min}$  and  $x_{max}$  indicate the initial value range of input variable, and  $p$  denotes the converted input variable. This process enhances the model's performance for diabetes prediction by scaling features to a uniform range while preserving relative relationships among data points for diabetes prediction. Then, the pre-processed input is fed into the data imbalance process to data balancing.

#### 3.3 Data imbalance

After pre-processing, the ADASYN balances the unbalanced data in diabetes prediction. ADASYN focuses on generating synthetic samples for a minority class in imbalanced datasets. By adaptively adjusting sampling depending on the density of local data, ADASYN generates a more representative training set which improves the accuracy of the predictive model for diabetes prediction. It performs a distribution of different minority class instances by weighting them according to the learning levels. The purpose of ADASYN [24] is to establish a distribution of weight for minority classes which represents the learning challenges during the model's training. Hence, synthetic data is presented in every minority class. Moreover, synthetic data is included for balancing classes which enable them to effectively learn data. There are 2 ways in that ADASYN technique is enhanced. It decreases bias due to an imbalance class and the boundary of classification decision is shifted across various sample data. A dataset  $D_{tr}$  is trained with samples  $m$

which has  $\{x_i, y_i\}, i = 1, \dots, m$ , where,  $x_i$  is the instance which has  $X$  feature space with dimension  $n$ . The class imbalance degree  $d$  is computed utilizing Eq. (2).

$$d = \frac{m_s}{m_l} \tag{2}$$

Where,  $d \in (0, 1]$ ,  $m_s$  indicates the majority classes, and  $m_l$  represents the minority classes. If  $d < d_{th}$  where,  $d_{th}$  denotes the threshold in an imbalanced class ratio. A maximum threshold value is a toleration for executing an imbalanced class ratio. Synthetic data generation  $G$  is represented by instances for minority class that is formulated in Eq. (3).

$$G = (m_l - m_s) \times \beta \tag{3}$$

Where,  $\beta \in [0, 1]$  is called a parameter to perform defined balanced levels which produce synthetic data. A  $\beta = 1$  balances a dataset that develops during generalization. For instance, if a minority class determines nearest neighbors utilizing Euclidean distance, then class computes a ratio which is expressed in Eq. (4).

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \tag{4}$$

Where,  $\hat{r}_i$  determines the density distribution. ADASYN utilizes the density distribution  $r_i$  to determine amount of samples in synthetic data. An imbalance is solved by including synthetic data for every minority weight, as well as synthetic data of minority weights computed by employing Eq. (5).

$$g_i = \hat{r}_i \times G \tag{5}$$

Where,  $r_i$  indicates the minority class weight and  $G$  indicates several synthetic samples that require a minority class that is expressed in Eq. (6). A minority class for every data  $x_i$  and synthetic data  $g_i$  are included for the selection of random minority data, and then generates the synthetic data. ADASYN not only presents data distribution as coefficient  $\beta$ , but also learns techniques by addressing challenges during the process of learning diabetes prediction. By performing this operation, ADASYN solves the imbalanced data issue effectively. After solving the data imbalance issue, the PSRNN is performed to predict diabetes effectively.

### 3.4 Prediction

The PSRNN is used for diabetes prediction after performing a data imbalance process. The RNN captures the temporal dependencies in patient data like the glucose level which generates context for more accurate predictions. The sequential processing enables them to include prior information which leads to dynamic predictions for evolving patient conditions. Also, RNN manages variable-length input sequences by using their internal memory or state which demonstrates the diverse time intervals among patient measurements in diabetes monitoring. During training, swish introduces a learnable parameter that adjusts the activation function's shape and enhances model performance. This adaptability enables swish to capture more complex data patterns which is beneficial in tasks such as diabetes prediction. Furthermore, parametric swish activation function helps to minimize vanishing gradient issue which provides more stable and effective training. Therefore, these process makes activation function to improve the predictive accuracy and robustness of neural networks applied to diabetes prediction process. Fig. 2 shows the architecture of RNN cell.

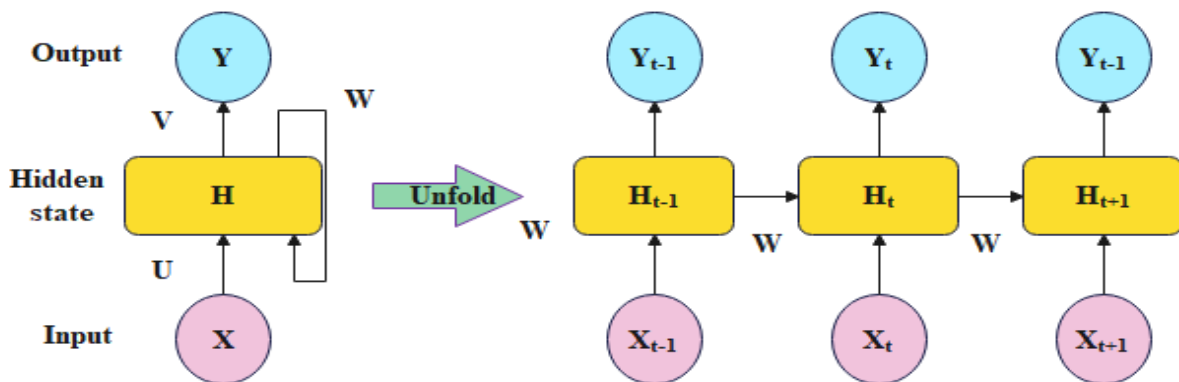


Figure. 2 The architecture of RNN Cell

The output from step  $n - 1$  is passed back into the network to generate the outcome of step  $n$  for each subsequent step. This RNN network is utilized for diabetes prediction role by forecasting the next value based on the past values. RNN has one hidden layer with its spread-out shape and the hidden state is represented as  $H_t$  based on the prior hidden state values and present time step-out output which is represented in Eq. (6).

$$H_t = \alpha_H (UX_t + WH_t + b_t) \quad (6)$$

Where,  $X_t$  represents the input data at time  $t$  and  $Y_t$  denotes the RNN output. The output state  $Y_t$  for input  $X_t$  is calculated based on hidden state  $H_t$  at time step  $t$  which is indicated in Eq. (7).

$$Y_t = \alpha_y (VH_t + b_y) \quad (7)$$

Here,  $U$ ,  $V$ ,  $W$ , and  $b$  indicate the parameter metrics and vectors.

The swish activation is a combination of linearity of ReLU and non-linearity of sigmoid that leads to improved learning dynamics. The choice of a fixed coefficient limits its effectiveness in capturing the complex data patterns. In order to address this issue, the parametric swish activation allows the model to learn an optimal value of the sigmoid coefficient which provides more flexibility and improves the

model's performance which is expressed in Eq. (8).

$$f(x) = \begin{cases} \alpha \cdot x \cdot \text{sigmoid}(\delta \cdot x), & \text{if } x \leq \beta \\ x, & \text{if } x > \beta \end{cases} \quad (8)$$

Where,  $f(x)$  represents the parametric swish activation function. The swish function is applied to the input  $x$  and the output is scaled by a learnable parameter  $\alpha$ . Also, the threshold parameter  $\beta$  and a slope parameter  $\delta$  control the behavior of the activation function. By allowing the model to learn the optimal sigmoid coefficient during training, PSRNN captures complex temporal patterns present in the diabetes-related data. This adaptability improves the data capacity and learning ability in neural networks which provides robust and accurate diabetes predictions.

Table 1 indicates a description of the notation table.

## 4. Results

The proposed PSRNN is simulated using Python 3.11 environment with Windows 10 operating system, 128 GB RAM, 1TB memory, 22 GB GPU, and Intel i9 processor. The performance metric of accuracy evaluates how often a model makes a correct prediction over an entire dataset, while precision is the ratio between TP and all positives, and Recall the measures correctly identified TP. Additionally, the F1-score defines a combination of precision and recall. The mathematical formula for these performance metrics is expressed in Eqs. (9) to (12).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

$$F1 - \text{score} = \frac{2TP}{2TP+FP+FN} \quad (12)$$

Where,  $TP$  indicates True Positive,  $FP$  indicates False positive,  $TN$  represents True Negative, and  $FN$  denotes False Negative respectively.

### 4.1 Performance evaluation

The performance analysis of PSRNN is evaluated in Tables 2 and 3. Table 2 indicates the performance

Table 1. Notation Table

Symbol	Description
$min$ and $max$	specified input variable range
$x_{min}$ and $x_{max}$	initial value range of input variable
$p$	converted input variable
$m_s$	majority classes
$m_l$	minority classes
$d_{th}$	a threshold in imbalanced class ratio
$\hat{r}_i$	density distribution
$G$	number of synthetic samples
$Y_t$	output state
$X_t$	input state
$H_t$	hidden state
$t$	time step
$U, V, W, \text{ and } b$	parameter metrics and vectors
$x$	input
$f(x)$	parametric swish activation function
$\alpha$	learnable parameter
$\beta$	threshold parameter
$\delta$	slope parameter
$TP$	True Positive
$FP$	False positive
$TN$	True Negative
$FN$	False Negative

Table 2. Performance evaluation of prediction using PIDD dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PS-CNN	89.20	88.24	88.15	87.65
PS-DNN	91.05	90.35	89.65	90.47
PS-LSTM	93.87	92.35	93.12	92.05
PS-RNN	99.81	99.12	98.72	97.98

Table 3. Different activation functions for prediction technique using PIDD

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ReLU-RNN	88.56	86.34	87.04	88.21
Mish-RNN	89.54	87.32	86.50	85.29
PAU-RNN	90.17	89.57	90.11	88.64
ELU-RNN	92.57	91.05	92.87	91.36
PELU-RNN	93.88	94.27	94.05	93.57
S-RNN	95.22	94.35	95.14	95.67
PS-RNN	99.81	99.12	98.72	97.98

of prediction on the PIDD dataset. The performance of Parametric Swish-Convolutional Neural Networks (PS-CNN), PS-Deep Neural Networks (PS-DNN), and PS-Long Short-Term Memory (PS-LSTM) are the existing techniques compared with PSRNN. When compared to these existing techniques, the proposed PSRNN achieves a high performance of 99.81% accuracy due to its effective handling of sequential data by including PSRNN that balances both gradient stability and non-linearity, thereby enhancing the convergence speed and learning ability of the neural network.

Table 3 represents a different activation function for the prediction technique using PIDD. The outcomes of Rectified Linear Unit-RNN (ReLU-RNN), Mish-RNN, Pade Activation Unit-RNN (PAU-RNN), Exponential Linear Unit-RNN (ELU-RNN), Parametric ELU-RNN (PELU-RNN), and Swish-RNN (S-RNN) are compared with PS-RNN. ReLU suffers from dying problem due to neurons are inactive, ELU and PELU causes vanishing gradient issue, Mish and PAU is computationally more intensive. When compared to these techniques, the PS-RNN achieves a superior accuracy of 99.81% due

to parametric swish providing a smoother gradient which improves the learning ability in RNN, as opposed to ELU/PELU negative slopes. The proposed PS activation function has more flexibility and enables better learning dynamics in diabetes prediction.

Fig. 3 illustrates a graphical representation of the prediction performances on the Germany dataset. The performance of PS-CNN, PS-DNN, and PS-LSTM are contrasted in relation to the proposed PS-RNN technique. The proposed PS-RNN achieves a high performance of 0.992 due to its effective handling of sequential data by adding a parametric swish that balances both non-linearity and gradient stability which in turn enhances the convergence speed and learning ability of the neural network.

Fig. 4 denotes a graphical representation of different activation techniques on the Germany dataset. The performance of ReLU-RNN, Mish-RNN, PAU-RNN, ELU-RNN, PELU-RNN, and S-RNN are compared with that of the PS-RNN. In relation to the existing techniques, the proposed approach attains a high accuracy of 0.992 because the parametric swish generates a smoother gradient which improves the learning ability in RNN.

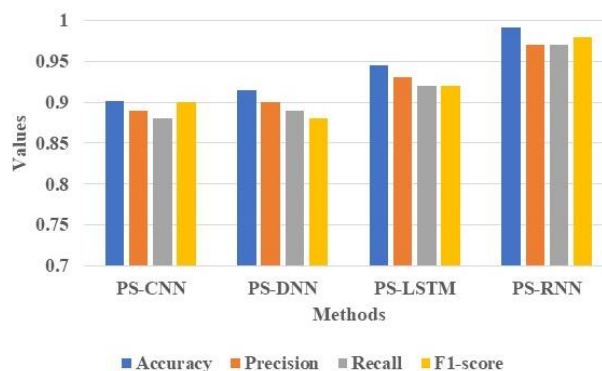


Figure. 3 Graphical representation of prediction performance using Germany dataset

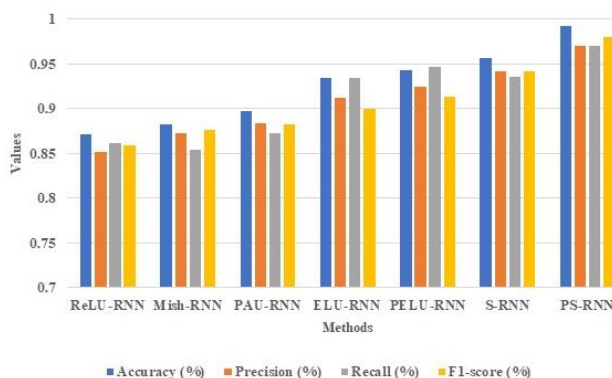


Figure. 4 Graphical representation of different activation functions using Germany dataset

Table 4. Comparative analysis with existing methods by utilizing PIDD dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
K-means clustering with SVM [16]	98.7	98.6	96.8	97.5
DCNN [17]	96.13	94.44	94.42	94.46
Statistical predictive [18]	85.9	N/A	N/A	N/A
Proposed PSRNN	99.81	99.12	98.72	97.98

Table 5. Comparative analysis with existing methods using Frankfurt Hospital, Germany dataset

Methods	Accuracy	Precision	Recall	F1-score
ANN [19]	0.958	0.95	0.95	0.95
RF [20]	0.989	N/A	0.95	0.972
Proposed PSRNN	0.992	0.97	0.97	0.98

## 4.2 Comparative analysis

Table 4 denotes a comparative analysis of the existing techniques on the PIDD dataset. The existing methods namely, K-means clustering with SVM [16], DCNN [17], and Statistical predictive approach [18] are compared with the proposed method.

Table 5 represents a comparative analysis of the existing methods on the Germany dataset. ANN [19] and RF [20] are the existing methods used for contrasting against the proposed method on the Germany dataset. The proposed PSRNN approach attains commendable results of 99.81% and 0.992 respectively on the PIDD and Germany datasets. Due to it effectively capturing the complex temporal patterns through adaptable non-linearity, it provides robust information propagation and gradient flow for sequential data that improves the overall performance for prediction.

## 4.3 Discussion

This section briefly discusses the limitations of the existing method and the advantages of the proposed method. The existing methods have limitations like K-means clustering's [16] dependency on Euclidean distance and consideration of isotropic clusters which make it less effective in capturing the complex distribution of data due to the clusters varying in size and shape. DCNN [17] suffers from class imbalance, leading to a biased model due to inadequate pattern learning which minimizes the

prediction performance. Furthermore, the Statistical approach [18] struggles with capturing complex interactions between variables because it is based on linear relationships, resulting in inaccurate prediction. RF [20] struggles with interpretability due to the complexity of multiple tree decisions. The proposed approach overcomes these existing methods' limitations by solving class imbalance issues and capturing complex patterns. By allowing the model to learn the optimal sigmoid coefficient during training, the PSRNN captures complex temporal patterns present in the diabetes-related data. This adaptability augments the data capacity and learning ability in neural networks, giving rise to robust and accurate diabetes predictions.

## 5. Conclusion

The PSRNN is proposed to effectively and accurately predict diabetes. RNN captures temporal data and manages variable-length input sequences by utilizing their internal memory which produces an accurate prediction. Moreover, Parametric swish activation allows the model to learn the optimal value of its coefficient which provides more flexibility and strengthens the model performance. Hence, data capacity and learning ability are increased during prediction. By performing this process, the PSRNN effectively predicts diabetes. As opposed to the existing techniques, SVM, DCNN, and statistical approach, the proposed PSRNN accomplishes a superior accuracy of 99.81% and 0.992 on the PIDD and Germany datasets, correspondingly. In the future, efficient DL techniques can be used to capture long-term dependency data for further enhancing the model's performances.

## Conflicts of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization, SKC and AJ; methodology, SKC; software, AJ; validation, AJ; formal analysis, SKC; investigation, AJ; resources, SKC; data curation, AJ; writing—original draft preparation, SKC; writing—review and editing, AJ; visualization, SKC; supervision, AJ; project administration, AJ.

## References

- [1] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques", *Healthcare Technology Letters*, Vol. 10, No. 1-2, pp. 1-10, 2023.

- [2] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective", *Computer Methods and Programs in Biomedicine*, Vol. 220, p. 106773, 2022.
- [3] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques", *Measurement: Sensors*, Vol. 25, p. 100605, 2023.
- [4] U. Ahmed, G. F. Issa, M. A. Khan, S. Aftab, M. F. Khan, R. A. T. Said, T. M. Ghazal, and M. Ahmad, "Prediction of diabetes empowered with fused machine learning", *IEEE Access*, Vol. 10, pp. 8529-8538, 2022.
- [5] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong, and F. H. Juwono, "Diabetes detection based on machine learning and deep learning approaches", *Multimedia Tools and Applications*, Vol. 83, No. 8, pp. 24153-24185, 2024.
- [6] C. Y. Chou, D. Y. Hsu, and C. H. Chou, "Predicting the onset of diabetes with machine learning methods", *Journal of Personalized Medicine*, Vol. 13, No. 3, p. 406, 2023.
- [7] A. Dođru, S. Buyrukođlu, and M. Ari, "A hybrid super ensemble learning model for the early-stage prediction of diabetes risk", *Medical & Biological Engineering & Computing*, Vol. 61, No. 3, pp. 785-797, 2023.
- [8] J. Abdollahi, and B. Nouri-Moghaddam, "Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction", *Iran Journal of Computer Science*, Vol. 5, No. 3, pp. 205-220, 2022.
- [9] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, "An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI", *Sensors*, Vol. 22, No. 19, p. 7268, 2022.
- [10] K. C. Howlader, M. S. Satu, M. A. Awal, M. R. Islam, S. M. S. Islam, J. M. W. Quinn, and M. A. Moni, "Machine learning models for classification and identification of significant attributes to detect type 2 diabetes", *Health Information Science and Systems*, Vol. 10, No. 1, p. 2, 2022.
- [11] S. Sadeghi, D. Khalili, A. Ramezankhani, M. A. Mansournia, and M. Parsaeian, "Diabetes mellitus risk prediction in the presence of class imbalance using flexible machine learning methods", *BMC Medical Informatics and Decision Making*, Vol. 22, No. 1, p. 36, 2022.
- [12] A. Nicolucci, L. Romeo, M. Bernardini, M. Vespasiani, M. C. Rossi, M. Petrelli, A. Ceriello, P. Di-Bartolo, E. Frontoni, and G. Vespasiani, "Prediction of complications of type 2 Diabetes: A Machine learning approach", *Diabetes Research and Clinical Practice*, Vol. 190, p. 110013, 2022.
- [13] M. Saberi-Karimian, A. Mansoori, M. M. Bajgiran, Z. S. Hosseini, A. Kiyoumarsioskouei, E. S. Rad, M. M. Zo, N. Y. Khorasani, M. Poudineh, S. Ghazizadeh, G. Ferns, H. Esmaily, M. and Ghayour-Mobarhan, "Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements", *Journal of Clinical Laboratory Analysis*, Vol. 37, No. 1, p. e24798, 2023.
- [14] C. J. Ejiyi, Z. Qin, J. Amos, M. B. Ejiyi, A. Nnani, T. U. Ejiyi, V. K. Agbesi, C. Diokpo, and C. Okpara, "A robust predictive diagnosis model for diabetes mellitus using Shapley-incorporated machine learning algorithms", *Healthcare Analytics*, Vol. 3, p. 100166, 2023.
- [15] S. A. Alex, N. Z. Jhanjhi, M. Humayun, A. O. Ibrahim, and A. W. Abulfaraj, "Deep LSTM model for diabetes prediction with class balancing by SMOTE", *Electronics*, Vol. 11, No. 17, p. 2737, 2022.
- [16] N. Arora, A. Singh, M. Z. N. Al-Dabagh, and S. K. Maitra, "A novel architecture for diabetes patients' prediction using K-means clustering and SVM", *Mathematical Problems in Engineering*, Vol. 2022, No. 1, p. 4815521, 2022.
- [17] K. K. Patro, J. P. Allam, U. Sanapala, C. K. Marpu, N. A. Samee, M. Alabdulhafith, and P. Plawiak, "An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques", *BMC Bioinformatics*, Vol. 24, No. 1, p. 372, 2023.
- [18] K. Azbeg, M. Boudhane, O. Ouchetto, and J. S. Andaloussi, "Diabetes emergency cases identification based on a statistical predictive model", *Journal of Big Data*, Vol. 9, No. 1, p. 31, 2022.
- [19] M. K. Gourisaria, G. Jee, G. M. Harshvardhan, V. Singh, P. K. Singh, and T. C. Workneh, "Data science appositeness in diabetes mellitus diagnosis for healthcare systems of developing nations", *IET Communications*, Vol. 16, No. 5, pp. 532-547, 2022.
- [20] M. O. Edeh, O. I. Khalaf, C. A. Tavera, S. Tayeb, S. Ghouali, G. M. Abdulsahib, N. E. Richard-Nnabu, and A. Louni, "A classification algorithm-based hybrid diabetes prediction model", *Frontiers in Public Health*, Vol. 10, p. 829519, 2022.



- [21] PIMA dataset link:  
<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>.
- [22] Frankfurt Hospital, Germany dataset link:  
<https://www.kaggle.com/code/linggarmaretva/frankfurt-hospital-diabetes-with-lgbmclassifier>
- [23] M. Shantal, Z. Othman, and A. A. Bakar, “A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization”, *Symmetry*, Vol. 15, No. 12, p. 2185, 2023.
- [24] A. Balaram, and S. Vasundra, “Prediction of software fault-prone classes using ensemble random forest with adaptive synthetic sampling algorithm”, *Automated Software Engineering*, Vol. 29, No. 1, p. 6, 2022.