scientific reports



OPEN

# An effective PO-RSNN and FZCIS based diabetes prediction and stroke analysis in the metaverse environment

M. Karpagam[1]✉, S. Sarumathi[2], A. Maheshwari[1], K. Vijayalakshmi[1], K. Jagadeesh[3], V. Bereznychenko[4]✉ & R. Narayanamoorthi[5]

**Chronic disease (CD) like diabetes and stroke impacts global healthcare extensively, and continuous monitoring and early detection are necessary for effective management. The Metaverse Environment (ME) has gained attention in the digital healthcare environment; yet, it lacks adequate support for disabled individuals, including deaf and dumb people, and also faces challenges in security, generalizability, and feature selection. To overcome these limitations, a novel probabilistic-centric optimized recurrent sechelliott neural network (PO-RSNN)-based diabetes prediction (DP) and Fuzzy Z-log-clipping inference system (FZCIS)-based severity level estimation in ME is carried out. The proposed system integrates Montwisted-Jaco curve cryptography (MJCC) for secured data transmission, Aransign-principal component analysis (A-PCA) for feature dimensionality reduction, and synthetic minority oversampling technique (SMOTE) to address data imbalance. The diagnosed results are securely stored in the BlockChain (BC) for enhanced privacy and traceability. The experimental validation demonstrated the superior performance of the proposed system by achieving 98.97% accuracy in DP and 98.89% accuracy in stroke analysis, outperforming existing classifiers. Also, the proposed MJCC technique attained 98.92% efficiency, surpassing the traditional encryption models. Thus, the proposed system produces a secure, scalable, and highly accurate DP and stroke analysis in ME. Further, the research will extend the approach to other CD like cancer and heart disease to improve the predictive performance.**

**Keywords** Chronic disease monitoring (CDM), Diabetes prediction (DP), Stroke analysis (SA), Wearable devices (WD), Internet of Things (IoT), Deep learning (DL), Metaverse environment (ME) in healthcare, Probabilistic-centric optimized recurrent, Sechelliott neural network (PO-RSNN), Fuzzy Z-log-clipping inference system (FZCIS)

**List of symbols**

| | |
|---|---|
| $\wp_n$ | Patient data |
| $r^{\Xi}$ | Range of random number |
| $N$ | Number of patients |
| $\phi^{pub}$ | Public key for encryption |
| $\left(w, w^2, w^8, q, q_4, q_6\right)$ | Affine points on the curve |
| $\aleph^{\nabla}$ | Generated curve |
| $\partial, \beta$ | Co-ordinates in the curve |

[1]Department of Computational Intelligence, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Chennai, Tamil Nadu 603203, India. [2]Department of Artificial Intelligence and Data Science, K.S.Rangasamy College of Technology, Tiruchengode, Tamil Nadu 637215, India. [3]Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai 600062, India. [4]Department of Theoretical Electrical Engineering and Diagnostics of Electrical Equipment, Institute of Electrodynamics, National Academy of Sciences of Ukraine, Beresteyskiy avenue, 56, Kyiv 03057, Ukraine. [5]Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu 603 203, India. ✉email: karpagam1@srmist.edu.in; vika.bereznichenko@i.ua

| | |
|---|---|
| $\left(\hbar^{1},\hbar^{2}\right)$ | Cipher text 1 and 2 |
| $\wp_{n}\left(\ell\right)$ | Original message's point on the curve |
| $\hbar_{g}$ | Encrypted message |
| $g=1,2,\ldots G$ | Number of encrypted messages |
| $\gamma_{t}$ | Collected data |
| $T$ | Number of collected data |
| $\left(\sigma^{mean}\right)$ | Mean value of the non-missing values |
| $\mathrm{K}_{l}^{\Theta}$ | Pre-processed data |
| $l=1,2,\ldots L$ | Number of pre-processed data |
| $\left(\tau^{nest}\right)$ | Minority instance |
| $\left(\rho^{new}\right)$ | New instance |
| $\Re^{ptr}$ | Random value |
| $\Upsilon_{b}$ | Balanced data |
| $b=1,2,\ldots B$ | Number of balanced data |
| $F_{o}$ | Extracted features |
| $O$ | Number of extracted features |
| $\upsilon^{sta}$ | Standardization process |
| $\omega_{mean}$ | Mean value |
| $\varpi_{std}$ | Standard deviation value |
| $Cov^{m}$ | Covariance matrix |
| $\left(F_{1},F_{2}\right)$ | Extracted features |
| $\xi_{eig}$ | Scalar value |
| $\psi^{v}$ | Non-zero vector |
| $\left(\alpha\rho^{\circ}\right)$ | Aransign function |
| $\eta_{a}$ | Dimensionality-reduced features |
| $A$ | Number of dimensionality-reduced features |
| $\exp$ | Exponential factor |
| $\delta\alpha f$ | Sechelliott activation function |
| $\left(\vartheta_{m}\right)$ | Weight parameter |
| $\left(\varsigma_{fit}\right)$ | Fitness function |
| $\left(\vartheta_{m}^{\circ}\right)$ | Rolling phase update position |
| $\left(\vartheta_{m-1}\right)$ | Weight parameter at the previous iteration |
| $n^{coeff}$ | Natural coefficient |
| $d^{coeff}$ | Deflection coefficient |
| $\Phi^{con}$ | Constant value |
| $\ell\varsigma$ | Changes in light intensity |
| $\chi^{\infty w}$ | Global worst position |
| $\left(\vartheta_{m}^{2\circ}\right)$ | Updated position regarding the dancing process |
| $\partial f$ | Deflection angle |
| $\left(\vartheta_{m}^{\Re\infty}\right)$ | Reproduction phase position update |
| $\phi^{\nabla}$ | Lower bound value |
| $\varphi^{\Delta}$ | Upper bound value |
| $\left(\vartheta_{m}^{\wp\circ}\right)$ | Foraging strategy updated position |
| $\mathrm{P}^{\Omega}$ | Probability factor |
| $\left(\vartheta_{m}^{\infty st}\right)$ | Stealing behavior updated position |
| $\vartheta_{m}^{\Theta f}$ | Neighborhood beetle with food |
| $\mathrm{O}^{\vartheta_{m}}$ | Optimized weight parameter |
| $\left(Hd\right)$ | Hidden Layer |
| $\iota^{bias}$ | Bias value |
| $\left(\lambda^{dia}\right)$ | Diabetic class |
| $\left(\upsilon_{non}\right)$ | Non-diabetic class |
| $\left(\mathrm{X}^{out}\right)$ | Classified diabetes result |
| $C^{gory}$ | Factor |
| $\left(\nu',\nu'',\nu''',\nu''''\right)$ | Factor range |
| $\left(\kappa_{risk}\right)$ | Risk score |
| $\left(Fuz^{\Theta}\right)$ | IF–THEN rule |
| $\mathrm{Z}^{set}$ | Fuzzy set |
| $\lambda^{fuzy}$ | Membership function |
| $\chi^{is}$ | Input space |
| $\left(c^{d}\right)$ | Crisp data |
| $\left(f^{d}\right)$ | Fuzzy data |
| $\left(b^{f}\right)$ | Fuzzification unit |
| $\left(a^{d}\right)$ | Defuzzification unit |
| $mn$ | Mean value in severity estimation |
| $sn$ | Standard deviation value in severity estimation |
| $\left(h^{\circ}\right)$ | High severity of diabetes |
| $\left(m^{\circ}\right)$ | Moderate severity of diabetes |
| $\left(l^{\circ}\right)$ | Low severity of diabetes |

| | |
|---|---|
| $Fuzzy^{res}$ | Fuzzy outcome |
| $(str^{data})$ | Information related to stroke patient |
| $(pre^{data})$ | Pre-processed stroke data |
| $(bal^{data})$ | Balanced stroke data |
| $(ext^{fea})$ | Extracted stroke features |
| $D_y$ | Dimensionality-reduced features |
| $Y$ | Number of dimensionality-reduced features |
| $(Cls_{out})$ | Stroke classified outcome |
| $stk$ | Stroke class |
| $non^{stk}$ | Non-stroke class |

Chronic illnesses have a major influence on global health and societies around the world. CD, such as diabetes and stroke, is among the leading causes of mortality worldwide, thus requiring continuous monitoring and early detection for better treatment[1]. CD is generally defined as a long-lasting disorder, which gradually increases over time and requires persistent healthcare treatment[2]. Likewise, the CD might be treated but not cured easily. Cancer, diabetes, stroke, obesity, and asthma are common chronic conditions that affect both adults and children[3]. But, chronic conditions like diabetes and stroke have been recently considered as the most dreadful disorders[4]. In general, diabetes is known as a metabolic condition categorized by increased levels of blood glucose and also affects the organs of the body, such as blood vessels and nerves[5]. If untreated, it leads to cardiovascular disease, nerve damage, and kidney failure[6]. Likewise, a neurological impairment led by a blockage and disturbance in the blood supply to the brain part is known as a stroke[7]. In fact, CD could potentially reduce the quality and health of the individual's life by causing immense pain and functional inability. Thus, to monitor individuals with CD, an automatic medical guidance system utilizing artificial intelligence (AI) is introduced[8].

With the evolution of ME-based healthcare, the existing works have utilized wearable IoT sensors and AI-based models for continuous monitoring[9]. ME has gained more attention in healthcare management, including CD Monitoring (CDM), with the quick development of Internet of Things (IoT) technology[10]. An ME is known as a virtual or digital platform in which humans can interact with objects in real-time through digital twin models[11]. Moreover, different immersive schemes like Augmented Reality (AR), Virtual Reality (VR), and Extended Reality (XR) are included in ME for ensuring an efficient user experience[12,13]. Wearable Devices (WDs) act as the most powerful tool in the ME to capture and transmit patients' medical data to the corresponding medical industry[14,15]. In addition, BC is integrated with ME to create a secure and decentralized data storage and monitoring system. To ensure secure data transmission, authentication schemes like Elliptic Curve Cryptography (ECC), Data Encryption Standard (DES), and K-anonymity are utilized[16]. But, they often fail to balance security and computational efficiency[17].

In recent times, medical experts have implemented various ME-centric CDM models using AI[18]. To predict the CD like diabetes and stroke, traditional works used AI approaches, namely Support Vector Machine (SVM), Artificial Neural Network (ANN), Fuzzy Inference System (FIS), Logistic Regression (LR), Recurrent Neural Network (RNN), and Deep Neural Network (DNN)[19]. But, some drawbacks like algorithm bias, variation in disease progression, and ethical concerns were present in the conventional systems. The existing models also struggled to generalize across different demographics with multi-datasets, thus limiting the real-world application regarding CD monitoring[20,21]. Also, prevailing work was not generalized to support disabled persons in the ME. Hence, this paper proposes a novel PO-RSNN and FZCIS-based DP and SA in the ME. Therefore, the CD of normal and disabled people in the ME is efficiently managed in the research framework.

Some limitations of the traditional models are given below:

- Existing models were not regularized well enough to provide communication support for disabled persons in ME-centric CD monitoring.
- In Ref.[22], the severity level of diabetes was calculated according to the unstable score values, which affected the model performance.
- Due to the insignificant consensus mechanisms and data authentication process, the technique in Ref.[23] had security and privacy issues.
- The model in Ref.[24] evaluated the stroke by only considering the upper body movement of the patients, thus increasing the misclassification rate.
- Owing to the imbalanced dataset and high dimensionality of the features, most of the prevailing works obtained limited outcomes and high complexity.

The research technique's significant contributions are explained further:

- The proposed work is proficiently designed to provide communication support for disabled persons in the ME.
- An effective FZCIS is introduced to estimate the severity level of diabetes according to the risk score.
- The Proof-of-Authority (PoA) protocol and MJCC technique are employed in the proposed system to ensure secure healthcare data transmission.
- To perform SA in the proposed work, the crucial biological and vital parameters of the patients are considered, which improves the model's reliability.
- SMOTE is used in the research framework to balance the dataset classes. Moreover, a novel A-PCA is established to minimize the features' dimensionality.

The novelty of the proposed system is given in detail as follows, The proposed PO-RSNN-based DP and FZCIS-based severity estimation framework has multiple novel contributions, overcoming the limitations of the existing works. This work introduces PO-RSNN, which integrates the Sechelliott Activation Function (SAF) to prevent the vanishing gradient problem, and Probabilistic Dung Beetle Optimization (P-DBO) is used to enhance the weight optimization, leading to reduced overfitting. Also, the FZCIS is proposed for severity estimation, which provides more stable and computationally efficient results. Unlike the existing works that failed to overlook data security, the MJCC with Proof-of-Authority (PoA) BC consensus mechanism is utilized in the proposed system to ensure the secured storage and privacy of the patient's records. Further, the model is designed with ME-based communication support for disabled individuals, such as deaf and dumb people. This research focuses on two diverse datasets, namely the Diabetes Prediction Dataset and the Stroke Prediction Dataset, incorporating cross-validation for better generalizability and adaptability. Thus, these innovations help to provide a highly accurate and secure AI-based healthcare framework, outperforming the traditional models in CD prediction and management.

The rest of the article is arranged as: The related survey is discussed in section "Literature survey", the proposed work is mathematically explained in section "Proposed methodology for PO-RSNN and FZCIS-based diabetes prediction and stroke analysis in metaverse environment", the performance analysis of the proposed model is demonstrated in section "Results and discussion", and the article is concluded with future direction in section "Conclusion".

## Literature survey

Reference[23] presented secure IoT with a BC-centric monitoring system for DP utilizing ML approaches. Initially, by using the IoT sensors, the risk factors were collected from the patients. Next, the risk factors were inputted into the Diabetes Mellitus Prediction Model (DMPM). In this, the Random Forest (RF) algorithm was utilized to predict the diabetes condition. Lastly, to ensure data privacy, the hash value of the diagnosed results was preserved and stored in the BC. However, this system was less secure owing to the traditional consensus protocol and authentication schemes[25]. established a 5G-centric diabetes management model utilizing AI. In this, to improve the efficiency of the data transmission, a delay-aware Resource Allocation (RA) optimization centered on a double-queue model was used. Then, a Deep Forest Algorithm (DFA) was established in the application layer to classify the collected data as normal or diabetes. This approach acquired higher accuracy. But, it had maximum access delay owing to a greater number of collisions in the network. Reference[26] explained user-cloud-centric ensemble framework for type-2 DP along with diet plan recommendations. This approach was assessed by using the Pima Indian Diabetes (PID) dataset. Primarily, the input data was subjected to the process of missing value imputation. Then, to perform diabetes detection, the ensemble models, including Decision Tree, SVM, and ANN, were employed. Moreover, based on the diagnosed outcomes, a diet plan was provided. This technique achieved superior outcomes and high scalability. Nevertheless, this technique had memory requirement issues and computational complexity.

Reference[24] introduced AI-centric smart post-stroke assessment utilizing wearable devices. Initially, by using the wearable sensors, the body segment's motion was recorded. In this, to predict the post-stroke of the individuals, a Multi-Level Meta Learner (MLML) ensemble classification model was introduced. Hence, the analysis outcomes displayed that the model obtained superior computational efficiency. Since this model considered only the upper body movement of the patients for SA, it obtained a high misclassification rate.

Reference[27] established a DP model utilizing blended ensemble learning approaches. Here, steps like data normalization, feature selection, and DP were comprised. To predict the collected data as diabetes or non-diabetes, the ensemble models named Bayesian networks and radial basis functions were used. This ensemble method attained a high accuracy rate. However, it had a high processing time owing to the base learners' blending process. Reference[28] presented a forecasting framework for disease progression in stroke patients utilizing digital twins ML models. In this, to create a digital twin model, a variational-autoencoder was employed. To predict the disease progression in ischemic stroke patients, the digital twin model was used. This approach helped to improve clinical decision-making and provided virtual arms for clinical trials. But, as only 244 individuals were considered after data processing and filtering, this approach was not generalized well enough to handle large data structures.

Reference[22] explored DL-centric diabetes mellitus prediction along with severity level estimation. Primarily, the input data was pre-processed by utilizing the Switching Midvalue-centric Morphological Filter (SMVMF). Next, the important features were extracted and then inputted to the Optimal Weighted Deep Artificial Neural Network (OWDANN). To classify the collected data as diabetes and non-diabetes, the OWDANN was used. Subsequently, to predict the severity level of the diabetes patients, the Great-circle Distance-based Hierarchical Clustering (GDHC) was used. As per the analysis outcomes, the model achieved high efficiency. But, this work estimated the severity level of diabetes based on unstable score values. Reference[29] propounded a prediction model for ischemic stroke recurrence utilizing DL approaches. Initially, to train the prediction model, the patient data was collected from the publically available resources. In this, to detect ischemic stroke disease, a Back Propagation (BP) network and Multivariate Logistic Regression (MLR) were established. This approach had higher supremacy for disorder prediction, including ischemic stroke. But, owing to the increased number of iterations, the BP network took considerable training time.

Reference[30] implemented a hybrid DL model for stroke prediction utilizing a mobile AI smart hospital platform. In this, to implement the stroke prediction model, the Electromyography (EMG) signal dataset was used. Moreover, to connect AI with healthcare, a stacked Convolutional Neural Network (CNN) was established. Next, to perform SA using EMG signal, models like the Group Handling Method (GMDH) and Long Short Term Memory (LSTM) were integrated. This technique attained higher accuracy in stroke prediction utilizing EMG. Nevertheless, due to the channel variation of the signal, the GMDH had memory requirement issues and computational complexity.

Reference[31] explained ML-centric diabetes healthcare disease prediction model. To assess the prediction model, the Pima Indian Diabetes Database (PIDD) was employed. In this work, key processes like data pre-processing and classification were included. Here, Logistic Regression (LR) was used for predicting the diabetes disorder. Hence, the experimental outcomes proved that the model obtained better outcomes with a minimum error rate. Still, owing to the random process of sampling distribution, this approach had poor hyper-parameter selection.

Reference[32] estimated DP model via the data mining techniques. Initially, the data related to diabetes was collected and preprocessed. Next, the Knowledge Discovery Dictionary (KDD) was used for the selection of features and interpretation of the data. Then, data mining techniques, such as RF, SVM, Logistic Regression, and Naive Bayes were used for the prediction of diabetes. Thus, the presence of diabetes was effectively predicted in this model. On the contrary, the missing values in the data reduced the performance of the model.

Reference[33] identified diabetes in the patient's data. Here, the patient data was collected and the important features were selected using the Boruta feature selection technique. Then, the K-Means + + technique was used for clustering the unsupervised data. Further, the ensemble classifier was utilized to identify diabetes. Hence, the diabetes was classified precisely. However, the data was not balanced, thus misleading the classification accuracy. Reference[34] determined DL clinical decision support system for the prediction of diabetes. The diabetes data was collected and pre-processed regarding data cleaning, normalization, and feature conversion. Then, the features were selected using the Extra Tree Classifier (ETC). Further, by using ANN, CNN, and LSTM, the DP was done. Thus, diabetes was identified with higher accuracy. Yet, the model was computationally complex and increased the DP.

Reference[35] envisaged the Ensemble of Light Gradient Boosting Machine (LGB) and Adaptive Boosting for diabetes identification. The data related to the diabetes of all age type patients were collected. Then, the features were extracted, and by utilizing the ensemble classifier, such as LGB and Adaptive Boosting, the presence of diabetes was predicted. Thus, the DP was done with low processing time. On the other hand, the severity of the CD was not analyzed, which reduced the decision-making. Reference[36] developed a Gait acceleration-based diabetes detection model. Initially, the data were collected from the wearable sensors of the patients. Then, the Gait acceleration was utilized to analyze the relationship between the features. Next, the hybrid DL model, such as CNN and LSTM (CNN-LSTM) was used to predict the diabetes from the data of the patient. Thus, the computational complexity was improved by the model. However, the model's misclassification rate was higher. Table 1 gives the summary of the related works.

## Proposed methodology for PO-RSNN and FZCIS-based diabetes prediction and stroke analysis in metaverse environment

This paper implements the PO-RSNN and FZCIS-based DP and SA by using the clinical data and vital parameters of the patients. Moreover, the research approach is strategically designed to provide support for disabled people who are unable to experience ME. Hence, Fig. 1 exhibits the proposed system's structural representation.

### Patient and doctor registration

Initially, the patients and doctors are registered in BC to access the ME. By using the MJCC technique, the public and private keys are generated during registration, which is explained in section "Data security". During registration, it is clear whether the patient is a normal or disabled person. Next, according to the nature of the individuals, the way of communication like speech-to-speech and text-to-speech or vice-versa is activated. This helps to the interaction of differently abled persons, such as deaf and dumb individuals. Then, the patient and doctor virtually communicate through the ME. The doctor collects the patient's healthcare behavior data via virtual conversation. At the same time, the biological and vital parameters like BMI, average blood glucose level, and HbA1c level are collected from the patients by utilizing IoT sensors. Hence, the collected patient data is signified as given in Eq. (1),

$$\wp_n \xrightarrow[data]{patient} \left\| \wp_1, \wp_2, \ldots\ldots \wp_N \right\|, \quad \text{Here, } n = 1, 2, \ldots N \tag{1}$$

where, $N$ is the number of patient data $\wp_n$.

### Data security

Next, to preserve the sensitive information of the patients, the patient data is encrypted. For this purpose, the proposed work uses the MJCC approach. The prevailing ECC is selected as it has fast encryption and decryption. It provides high security with a smaller key size, making it highly efficient in a resource-constrained environment. But, high computational complexity is caused by the negative point on the curve, making the computation more complex. Thus, for enhancing the system's security level, the proposed system establishes the Montwisted-Jaco (MJ) curve. Therefore, the proposed MJCC is derived further,

Primarily, the key generation is done for creating the public key and private key, which perform encryption and decryption, correspondingly. Moreover, the private key is selected from the random number that falls between the ranges $[1, N-1]$. In this, the public key $\left(\phi^{pub}\right)$ is created by applying the following expression,

$$\phi^{pub} = r^{\Xi} \cdot (w, q)\, ; \;\; r^{\Xi}\, [1, N-1] \tag{2}$$

| Author | Objective | Methods | Advantages | Drawbacks |
|---|---|---|---|---|
| (Hennebelle et al. 2024)[23] | DP via secured IOT with BC monitoring | Diabetes mellitus prediction model (DMPM) and Random Forest (RF) | Ensured data privacy by storing the diagnosed result in the BC | As the traditional consensus protocol and authentication scheme were utilized, the security was reduced |
| (R. Huang et al. 2022)[25] | Diabetes management model | Delay-aware resource allocation (RA) and deep forest algorithm (DFA) | Effectively transmitted the data and attained high diabetes classification accuracy | The network collision was not avoided, leading to higher access delays |
| (Prabhakar et al. 2024)[26] | Type 2 DP and diet plan recommendation | Decision Tree, support vector machine (SVM), and artificial neural network (ANN) | Attained higher scalability and prediction accuracy | Required higher memory and was computationally complex |
| (Razfar et al. 2023)[24] | Smart post-stroke assessment in the wearable device | Multi-Level Meta Learner (MLML) ensemble classifier | Computational efficiency was better | As only the upper body movement was concentrated, the misclassification rate was higher |
| (Mahesh et al. 2022)[27] | DP model | Bayesian networks and radial basis functions | DP accuracy was higher | The blending of base learners caused higher processing time |
| (Allen et al. 2021)[28] | Forecasting disease progression in stroke patients | Digital twin and variational-autoencoder | Improved clinical decision-making | The dataset was small and the generalization was reduced |
| (Annamalai & Nedunchelian 2021)[22] | Diabetes mellitus prediction and severity estimation | Switching midvalue-centric morphological filter (SMVMF) and optimal weighted Deep artificial neural network (OWDANN) | Diabetes was classified with high efficiency | The severity level estimation was poor due to the unstable score values |
| (Lu & Wang 2022)[29] | Ischemic stroke prediction | Back propagation (BP) network and multivariate logistic regression (MLR) | The ischemic stroke was detected with high accuracy | As the number of iterations increased, the training time also increased |
| (Elbagoury et al. 2023)[30] | Stroke prediction via smart hospital platform | Convolutional neural network (CNN), group handling method (GMDH), and long short term memory (LSTM) | Obtained higher stroke prediction accuracy | The variation of the signal channel led to more memory usage and computational complexity |
| (Krishnamoorthi et al. 2023)[31] | Diabetes healthcare disease prediction | Logistic regression (LR) | The error rate was minimized | The hyper-parameter selection was degraded due to the random sample distribution |
| (Rastogi & Bansal 2023)[32] | DP model | KDD, RF, SV, logistic regression, and Naive Bayes | Improved accuracy in DP | Missing values in the data reduced the performance |
| (Zhou et al. 2023)[33] | Diabetes identification model | Boruta, K-Means + +, and ensemble classifier | DP precision was higher | The absence of data imbalance led to improper classified output |
| (Al Reshan et al. 2024)[34] | DL clinical decision-making regarding DP | ETC, ANN, CNN, and LSTM | Higher accuracy in DP | The model was complex and slowed the DP |
| (Sai et al. 2023)[35] | Prediction of diabetes | LGB and adaptive boosting | Attained higher processing time | Reduction in proper decision-making |
| (Chee et al. 2024)[36] | Detection of diabetes | CNN + LSTM | Lowered computational complexity | The misclassification rate was higher |

**Table 1.** Summary of related works.

where, $(w, q)$ signifies the affine points on the curve. The public key is created based on the range and the affine points in the curve. Likewise, the proposed work establishes the MJ curve that elevates the framework's security performance. Hence, the proposed MJ curve is formulated as,

$$\aleph^{\nabla} : \left(\beta q^2\right) \rightarrow \left| w^8 + \partial w^2 + 2qw^2 + q_2 w^2 + q_4 w + q_6 + 1 \right| \tag{3}$$

Here, $\partial, \beta$ are the coordinates of the field $\aleph^{\nabla}$. After generating $\aleph^{\nabla}$, the patient data $\wp_n$ is encrypted by using the two cipher text, such as cipher text 1 $\left(\hbar^1\right)$, as provided in Eq. (4), and cipher text 2 $\left(\hbar^2\right)$, which is generated as per Eq. (5). The cipher text is produced as follows,

$$\hbar^1 = r^{\Xi} \cdot (w.q)\, \aleph^{\nabla} \tag{4}$$

$$\hbar^2 = \wp_n\,(\ell) + r^{\Xi} * \phi^{pub} \tag{5}$$

where, $\wp_n\,(\ell)$ is the original message's point on the curve. Lastly, the decryption process is performed with respect to $\left(\hbar^1\right)$ and $\left(\hbar^2\right)$ as,

$$\wp_n = \hbar^2 - r^{\Xi} \times \hbar^1 \tag{6}$$

Hence, the encrypted message is mathematically expressed in Eq. (7) as,

$$\hbar_g = (\hbar_1, \hbar_2, \ldots\ldots \hbar_G) \tag{7}$$

Here, $g = 1, 2, \ldots G$ specifies the number of encrypted messages $\hbar_g$. Subsequently, the preserved message is stored in the BC. In the proposed framework, the function of the BC is controlled by utilizing the consensus protocol, such as PoA. Hence, efficient and advanced security of the network can be ensured by the PoA protocol.

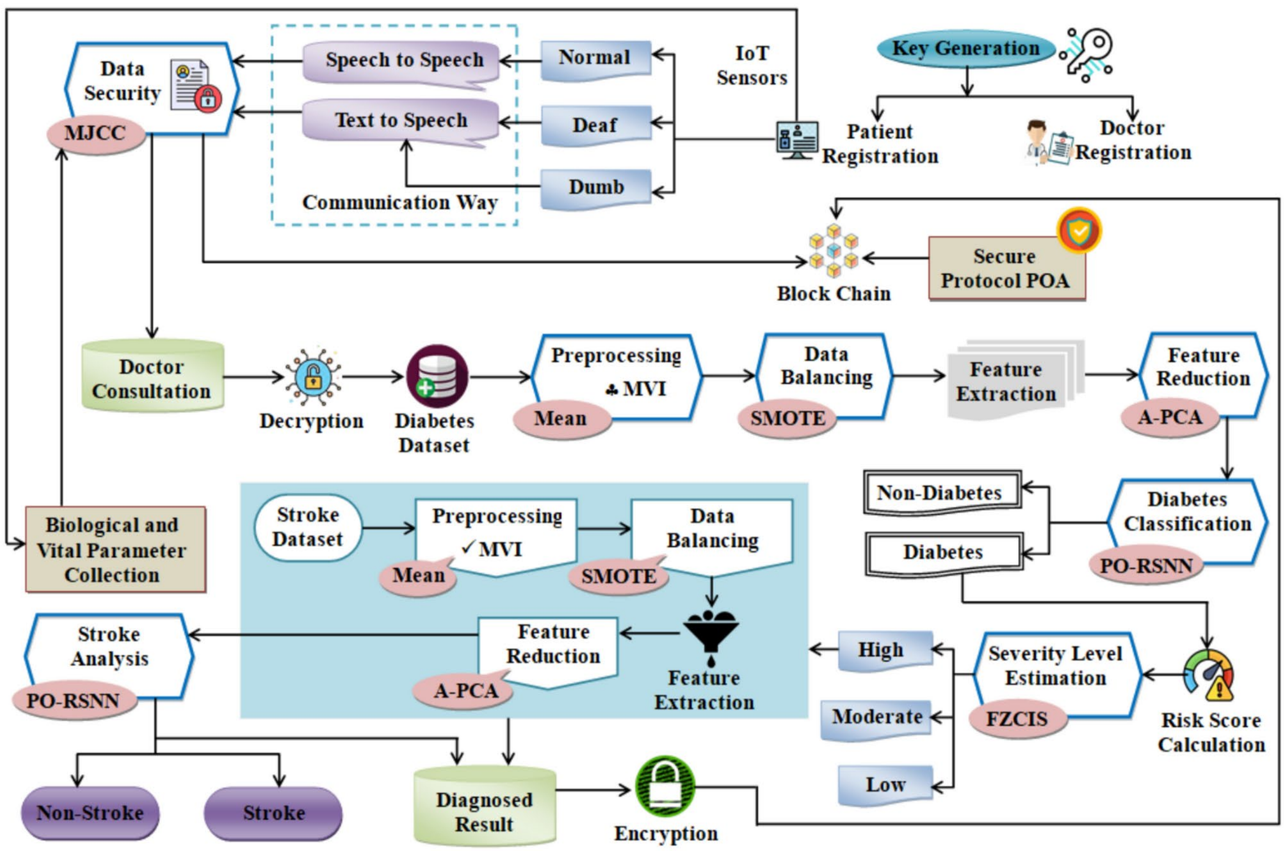



**Fig. 1**. Block diagram of the proposed system.

## Doctor consultation

Then, the encrypted message $\hbar_g$ is given to the doctor consultation phase, where the doctor decrypts the data utilizing the private key for performing the diagnosis process. Here, the pre-trained DP model is used to predict whether the patient is diagnosed with diabetes or non-diabetes. Hence, the DP model is discussed further,

*Diabetes dataset*

Initially, to train the diagnosis framework, the clinical data of the diabetes patients is gathered from the publically available resources. Therefore, the collected diabetes dataset is signified as,

$$\gamma_t = \{\gamma_1, \gamma_2, \ldots\ldots \gamma_T\}, \quad \text{Where,} \ \ t = 1\, to\, T \tag{8}$$

Where, $T$ indicates the number of collected data $\gamma_t$.

*Pre-processing*

Next, to upgrade the data quality and classification accuracy, the $\gamma_t$ is pre-processed. Here, the process of Missing Value Imputation (MVI) is carried out to replace the missing values in the dataset with the mean value $(\sigma^{mean})$ of the non-missing values in the dataset. The value $(\sigma^{mean})$ is given in Eq. (9). The process of MVI is described as shown in Eq. (10),

$$\sigma^{mean}(\gamma_t) = \left(\frac{sum\,|\gamma_t|}{T}\right) \tag{9}$$

$$\mathrm{K}_l^{\Theta} = \sigma^{mean}(\gamma_t) \overset{replace}{\longrightarrow} \sum_{l=1}^{L} \left\langle \mathrm{K}_l^{\Theta} \right\rangle \tag{10}$$

Here, $l = 1, 2, \ldots L$ depicts the number of pre-processed data $\mathrm{K}_l^{\Theta}$.

*Data balancing*

Next, the $\mathrm{K}_l^{\Theta}$ is balanced utilizing the Synthetic Minority Oversampling Technique (SMOTE). Moreover, the data balancing can upgrade the model's consistency by reducing the data shortage. The SMOTE is more suitable to produce additional data from the minority class. In the SMOTE, the minority class and majority class are selected from the $\mathrm{K}_l^{\Theta}$. Subsequently, the minority instance $\left(\tau^{nest}\right)$ is randomly assumed from the minority class. Then, the nearest neighbor is chosen from the minority instance. Likewise, the new instance is created by selecting the random neighbor $(ran^{\tau})$ among the nearest neighbor. Finally, for the minority class, the new instance is produced in the dataset. Hence, the new instance $(\rho^{new})$ is generated as given in Eq. (11).

$$\rho^{new} = \tau^{nest} + \left(ran^{\tau} - \tau^{nest}\right) \cdot \Re^{ptr} \tag{11}$$

$$\Upsilon_b = \rho^{new} \cdot |\Upsilon_1, \Upsilon_2, \ldots\ldots \Upsilon_B| \tag{12}$$

where, $\Re^{ptr}$ is the random value between 0 and 1 and $b = 1, 2, \ldots B$ signifies the number of balanced data $\Upsilon_b$ , which is estimated as given in Eq. (12).

*Feature extraction*
Here, the essential features, namely age, gender, hypertension, smoking history, heart disease, BMI, blood glucose level, and HbA1c level are extracted from the $\Upsilon_b$. Thus, the extracted features are defined as below,

$$F_o = (F_1, F_2, \ldots\ldots F_O) \quad \text{Where,} \;\; o = 1, 2, \ldots O \tag{13}$$

Where, $O$ depicts the number of extracted features $F_o$.

*Dimensionality reduction*
Subsequent to feature extraction, the dimensionality of the extracted features $F_o$ is reduced by utilizing the proposed A-PCA approach. The dimensionality reduction process aids in decreasing the proportionality of the features while preserving significant information. The prevailing PCA is chosen since it is effective and yields significant uncorrelated features. But, the PCA achieved high computational complexity owing to the process of eigenvalue approximation based on power iterative computation. For overcoming this problem, the proposed work introduces the Aransign function to approximate the eigenvalue. Therefore, the A-PCA process is derived further,

Primarily, the standardization process is performed for $F_o$. Here, each feature is individually analyzed to have a mean value of 0 and a standard deviation of 1. The standardization process is formulated as,

$$\upsilon^{sta} \to \left\| \frac{(F_o - \omega_{mean})}{\varpi_{std}} \right\| \tag{14}$$

Here, $\omega_{mean}$ and $\varpi_{std}$ are the mean and standard deviation values, correspondingly. These values are utilized for generating the standardization output $\upsilon^{sta}$. Next, the Covariance Matrix (CM) is generated to calculate the difference between the standardized features, such as $(F_1, F_2)$. In addition, the CM is utilized for determining the strength of the relationship among the independent features, which is expressed below,

$$Cov^m = \upsilon^{sta} \cdot \frac{1}{O-1} \cdot \sum_{o=1}^{O} \begin{bmatrix} (F_1, F_1) & (F_1, F_2) \\ (F_2, F_1) & (F_2, F_2) \end{bmatrix} \tag{15}$$

Likewise, the eigenvectors and eigenvalues are estimated from the covariance matrix $Cov^m$, which is expressed as,

$$Cov^m * \psi^v = \xi_{eig} \cdot \psi^v \tag{16}$$

Here, $\xi_{eig}$ is the scalar value, and $\psi^v$ is the non-zero vector. Furthermore, the scalar value and non-zero vector are assumed as the eigenvalue and eigenvector of the covariance matrix, correspondingly.

Next, the eigenvalue is approximated by using the Aransign function $(\alpha\rho^{\circ})$; thus, the computational efficiency of the model is improved. The eigenvalue is approximated regarding the eigenvalues and the exponential factor as,

$$\alpha\rho^{\circ}\left(\xi_{eig}\right) \to \frac{1 - \left(1 + 2\exp^{\xi_{eig}}\right)^{-1/2} * \xi_{eig}}{2\left(1 + |\xi_{eig}|\right)} \tag{17}$$

Subsequently, the principal component is selected by considering the eigenvector with the highest eigenvalue. The dimensionality of the features is constantly reduced by using the principal component. Hence, the dimensionality-reduced features are defined below,

$$\eta_a = \langle \eta_1, \eta_2, \ldots\ldots \eta_A \rangle, \;\; \text{Where}\, a = 1, 2, \ldots A \tag{18}$$

Where, $A$ is the number of dimensionality-reduced features $\eta_a$.

*Diabetes classification*
Here, the $\eta_a$ is given as input to the proposed PO-RSNN classifier that predicts whether the patient has diabetes disorder or not. The prevailing RNN is chosen since it is more efficient to handle healthcare behavioral data and

sequential information. But, it has vanishing gradient problems, which limits the classifier efficiency. In addition, it possesses overfitting issues owing to the random weight initialization. Overfitting is an issue in DL models, where the model learns patterns that are too specific to the training data, reducing its ability to generalize to new, unseen data. Thus, to enhance the learning efficiency of the neuron, the proposed work introduces the Sechelliott activation function (SAF). This regulates the neuron learning efficiency and prevents the overfitting caused by the traditional activation functions. Likewise, to optimize the weight parameter, the P-DBO technique is utilized. Dung Beetle optimizer (DBO) is selected since it produces a high convergence rate within less iteration. However, it is less efficient since it performs the foraging behavior of the dung beetles regarding the lower and upper-bound variables. Therefore, to perform a foraging strategy that elevates the significance of the system, the research framework employs the probabilistic distribution. Figure 2 presents the proposed PO-RSNN network diagram.

Hence, the proposed PO-RSNN is briefly described as,

Sechelliott activation function  In the proposed work, an effective activation named Sechelliott is employed, which improves the learning method of the neurons. The SAF is determined as,

$$\delta\alpha f \rightarrow \frac{1}{2} + \frac{0.5\eta_a \cdot 2}{1 + |\eta_a| * (\exp^{\eta_a} + \exp^{-\eta_a})} \tag{19}$$

Here, $\delta\alpha f$ depicts the Sechelliott activation function.

Input layer  Here, the input $\eta_a$ is collected and then transferred to the hidden layer, which processes the input by sharing the weight and bias value.

Weight initialization  In this layer, the weight parameter $(\vartheta_m)$ is optimized by using the proposed P-DBO algorithm. Here, the weight value is regarded as the member (dung beetle) of the population. In addition, the DBO is known as a meta-heuristic algorithm inspired by the biological behavior of the Dung Beetles (DB). Initially, the population's position is initialized in the local search space. Next, by considering the maximum classification accuracy, the fitness $(\varsigma_{fit})$ is calculated. The individual with superior fitness value is referred to as the best candidate solution. Then, the following search process updates the member's position, fitness value, and candidate solution. The searching process includes 5 phases, such as rolling, dancing, reproduction, foraging, and stealing, which are described below:

Step 1: (rolling)  The rolling process is carried out by considering the dung ball rolling in a straight line. Hence, the position of the DB is updated in the rolling phase $(\vartheta_m^{\circ})$, as given in Eq. (20).

$$\vartheta_m^{\circ} \rightarrow \vartheta_m + n^{coeff} * d^{coeff} \times \vartheta_{m-1} + \Phi^{con} \cdot \ell\varsigma \tag{20}$$

$$\ell\varsigma = \langle \vartheta_m - \chi^{\infty w} \rangle \tag{21}$$

where $\vartheta_{m-1}$ is the weight parameter at the previous iteration, $n^{coeff}$ and $d^{coeff}$ are the natural and deflection coefficients, correspondingly, $\Phi^{con}$ is the constant value, $\ell\varsigma$ is the changes in light intensity and is described in Eq. (21), and $\chi^{\infty w}$ is the global worst position.

Step 2: (dancing)  In this stage, the DB dances and then identifies an optimal path during obstacles. In the dancing process $\left(\vartheta_m^{2^{\circ}}\right)$, the position of the DB is updated as,

$$\vartheta_m^{2^{\circ}} = \vartheta_m + \|\tan \cdot \partial f * (\vartheta_m - \vartheta_{m-1})\| \tag{22}$$

Here, $\partial f$ is the deflection angle.

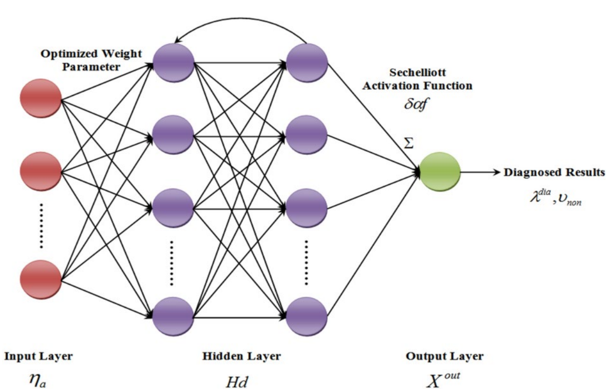



**Fig. 2**. The structural representation of the proposed PO-RSNN.

Step 3: (reproduction) Here, the female DBs' spawning location is chosen centered on the boundary selection strategy. Hence, the position of the DB updated in the reproduction phase $\left(\vartheta_m^{\Re\infty}\right)$ is depicted as,

$$\vartheta_m^{\Re\infty} = \left\|\vartheta_m - \phi^{\nabla}\right\| + \left\|\vartheta_m - \varphi_{\Delta}\right\| \tag{23}$$

where $\phi^{\nabla}$ and $\varphi_{\Delta}$ are the lower and upper bound, correspondingly. According to the position of the female DB, the current position $\left(\vartheta_m^{\Re\infty}\right)$ is updated.

Step 4: (foraging) To perform the foraging strategy of the DB, the proposed framework employs a probabilistic distribution function. Therefore, the position of the DB is updated in the foraging strategy $\left(\vartheta_m^{\wp^{\circ}}\right)$, which is displayed as,

$$\vartheta_m^{\wp^{\circ}} = \vartheta_m + \mathrm{P}^{\Omega} \cdot \left(\vartheta_m \cdot \phi^{\nabla} \leq \vartheta_m \cdot \varphi_{\Delta}\right) \tag{24}$$

where $\mathrm{P}^{\Omega}$ is the probability factor.

Step 5: (Thief) Here, the DB steals neighborhood beetles' food and then moves toward their location in the search space. Hence, the DB's position is updated regarding the stealing behavior $\left(\vartheta_m^{\infty st}\right)$, which is signified below,

$$\vartheta_m^{\infty st} = \vartheta_m^{\Theta f} + \left(\Phi^{con} \times \vartheta_m - \vartheta_m^{\Theta f}\right) \tag{25}$$

where $\vartheta_m^{\Theta f}$ is the neighborhood beetle with food. Next, the above-mentioned steps are continued until they converge. Hence, the optimized weight parameter is defined as $\mathrm{O}^{\vartheta_m}$.

Hidden layer This layer grasps the input from the input layer and then executes a computation process to give prediction results. The function of the hidden layer $(Hd)$ is formulated as,

$$Hd \rightarrow \delta\alpha f \times \left|\eta_a \cdot \mathrm{O}^{\vartheta_m}\right| + \iota_{bias} \tag{26}$$

Here, $\iota_{bias}$ is the bias value. The weight and bias values are added to the input and then activated to produce the final output.

Output layer Lastly, the output layer predicts whether the patients are diabetic $\left(\lambda^{dia}\right)$ or non-diabetic $(\upsilon_{non})$. The outcome of the PO-RSNN $\left(\mathrm{X}^{out}\right)$ is illustrated below,

$$\mathrm{X}^{out} = \left\{\lambda^{dia}, \upsilon_{non}\right\} \tag{27}$$

The pseudocode of the proposed PO-RSNN is given further,

**Input:** Dimensionality reduced features $\eta_a$.
**Output:** Classified results $\mathrm{X}^{out}$

**Begin**

**Initialize** $\delta\alpha f$, $\eta_a$, $\vartheta_m^{\Theta f}$, $\mathrm{O}^{\vartheta_m}$, and $Hd$

**For** 1 to $A$ number of $\eta_a$ do,

**Perform** input layer

**Apply** sechelliott activation function $\delta\alpha f$

**Optimize** weight parameter $\mathrm{O}^{\vartheta_m}$ using P-DBO

**Implement** hidden layer,

$$Hd \rightarrow \delta\alpha f \times \left|\eta_a \cdot \mathrm{O}^{\vartheta_m}\right| + \iota_{bias}$$

**Execute** output layer $\mathrm{X}^{out}$

**End For**

**Return** $\mathrm{X}^{out} = \left\{\lambda^{dia}, \upsilon_{non}\right\}$

**End**

The diabetes individuals are effectively classified by the proposed PO-RSNN, which enhances the lifecycle of the patients.

### Risk score calculation

In this, based on the factors like abdominal obesity, age, physical activity, and family history of diabetes, the risk score is calculated for $\lambda^{dia}$. Each factor is categorized into several cases, and each case has its own score value. For each category, the score values are provided based on the threshold value, which is displayed below,

$$C^{gory} \to \begin{cases} < \nu', & 0 \\ \nu'' \; to \, \nu''' & 20 \\ > \nu'''' & 30 \end{cases} \tag{28}$$

Here, $C^{gory}$ signifies the factor, and $(\nu', \nu'', \nu''', \nu'''')$ indicates the factor ranges. Hence, according to the factor value, the risk score is estimated. Hence, patients' risk score is depicted as $(\kappa_{risk})$.

### Severity level estimation

Likewise, by using the FSCIS technique, the severity level of the diabetes patients is estimated based on their $\kappa_{risk}$. The Fuzzy Inference System (FIS) is chosen since it provides efficient reasoning and accurate prediction. The FIS has the ability to handle imprecise, uncertain, and non-linear medical data more effectively. It interprets complex data, analyzes the relation between the data, and makes better decision-making. But, it has downsides like complex distribution of data because it utilizes the min–max algorithm to transform the crisp data to fuzzy data during defuzzification, which results in high complexity. Thus, to convert the fuzzy data into its original form, the proposed work establishes the z-log-clipping normalization. The proposed FZCIS is derived below,

Initially, by using the IF and THEN components, the decision rules are generated. The proposed fuzzy IF–THEN rules $\left(Fuz^{\Theta}\right)$ are framed as,

$$Fuz^{\Theta} = \begin{cases} IF\left(\kappa_{risk} \geq 60\right), & High \\ IF\left(\kappa_{risk} == [30, 50]\right), & Moderate \\ IF\left(\kappa_{risk} < 30\right), & Low \end{cases} \tag{29}$$

If the risk score of the patients is greater than or equal to 60, then it is assumed as high severity. Likewise, if the value of patients' risk score is between 30 and 50, then it is considered as a moderate level. Moreover, if the risk score is less than 30, then it is assumed as low severity. Also, the fuzzy membership function is used to map the fuzzification and defuzzification outcomes, exhibiting the fuzziness of the fuzzy system.

$$\mathrm{Z}^{set} \to \left\{\lambda^{dia}\left(\kappa_{risk}\right), \bar{\lambda}^{fuzy} \times \left(\lambda^{dia}\left(\kappa_{risk}\right) | \lambda^{dia}\left(\kappa_{risk}\right) \in \chi^{is}\right)\right\} \tag{30}$$

Here, $\bar{\lambda}^{fuzy}$ is the membership function, $\mathrm{Z}^{set}$ is the fuzzy set, and $\chi^{is}$ is the input space. The fuzzy set is evaluated based on $\bar{\lambda}^{fuzy}$, $\chi^{is}$, and $\kappa_{risk}$.

Then, via decision-making operators, the fuzzy operations are performed in the decision-making unit. Also, the crisp data $\left(c^{d}\right)$ is converted into fuzzy data $\left(f^{d}\right)$ in the fuzzification unit $\left(b^{f}\right)$, which is displayed as,

$$b^{f} = c^{d} \to f^{d} \tag{31}$$

Next, to perform a defuzzification unit $\left(a^{d}\right)$, the proposed work employs a z-log-clipping normalization. Here, the fuzzy data is transformed into crisp data regarding the IF–THEN condition of the mean and standard deviation values, which is determined as,

$$a^{d} = \begin{cases} IF\left(f^{d} > \max\left(f^{d}\right)\right), & THEN\, c^{d} = \max\left(\frac{\log\left(f^{d} - mn\right)}{sn}\right) \\ IF\left(f^{d} < \min\left(f^{d}\right)\right), & THEN\, c^{d} = \min\left(\frac{\log\left(f^{d} - mn\right)}{sn}\right) \end{cases} \tag{32}$$

where, $mn$ and $sn$ are the mean and standard deviation values, correspondingly. Lastly, the proposed FZCIS significantly estimates the severity of diabetes as high $(h^{\circ})$, moderate $(m^{\circ})$, and low $(l^{\circ})$, which is defined as,

$$Fuzzy^{res} \to \{h^{\circ}, m^{\circ}, l^{\circ}\} \tag{33}$$

Here, $Fuzzy^{res}$ is the fuzzy outcome. The proposed FZCIS's pseudocode is illustrated as,

**Input:** Risk score $\kappa_{risk}$
**Output:** Severity Level $Fuzzy^{res}$

**Begin**

**Initialize** $\lambda^{fuzy}$, $Fuz^{\Theta}$, $Z^{set}$, and $\chi^{is}$
**For** every features in the dataset do,
**Generate** fuzzy rules $\left(Fuz^{\Theta}\right)$
$IF\left(\kappa_{risk} \geq 60\right)$
$High$
$Else\,IF\left(\kappa_{risk} == [30,50]\right)$
$Moderate$
$Else\left(\kappa_{risk} < 30\right),$
$Low$
$End\ IF$
**Implement** membership function,
$Z^{set} \rightarrow \left\{\lambda^{dia}\left(\kappa_{risk}\right), \lambda^{fuzy} \times \left(\lambda^{dia}\left(\kappa_{risk}\right) \mid \lambda^{dia}\left(\kappa_{risk}\right) \in \chi^{is}\right)\right\}$
**Perform** decision-making unit
**Convert** $b^{f} = c^{d} \rightarrow f^{d}$
**Apply** z-log-clipping normalization,
**Evaluate** $\left(a^{d}\right)$
**End For**
**Return** $Fuzzy^{res} \rightarrow \left\{h^{\circ}, m^{\circ}, l^{\circ}\right\}$
**End**

Finally, the proposed FZCIS efficiently estimates the severity level of the diabetes individuals.

## Stroke analysis framework

Here, the diabetes persons with high severity $h^{\circ}$ are subjected to the SA process. In general, individuals, after being diagnosed with diabetes, face more complications, especially persons with high severity. Thus, SA is performed for the high severity diabetes patients to ensure people's health. The pre-trained SA framework is discussed further,

*Stroke dataset*
The significant information related to stroke patients is collected from the publically available resources, which are defined as $\left(str^{data}\right)$.

*Pre-processing*
Next, the $str^{data}$ is pre-processed under the mean value-based missing value imputation technique, which is already discussed in section "Pre-processing". Therefore, the pre-processed data is signified as $\left(pre^{data}\right)$.

*Data balancing*
Subsequently, the $pre^{data}$ is subjected to data balancing, which improves the classifier accuracy. For balancing the classes in the dataset, the proposed work uses the SMOTE technique, as previously explained in section "Data balancing". Next, the balanced data is indicated as $\left(bal^{data}\right)$.

*Feature extraction*
In this, features like id, gender, hypertension, age, heart disease, marital status, residence type, BMI, average glucose level, and smoking status are extracted from the $bal^{data}$. The extracted features are depicted as $\left(ext^{fea}\right)$

. Further, the dataset features for both datasets are described in Table 2.

The combination of these features in the DP and stroke analysis enables the model to capture physiological risk and behavioral risks. Feature engineering is utilized to evaluate the risk score, and this helps in the assessment of the severity of diabetes. Also, the selection of these stroke indicators gives accurate stroke analysis.

*Dimensionality reduction*
Additionally, the extracted features $ext^{fea}$ have high dimensionality, which mostly affects the model's efficiency. Hence, the dimensionality of the $ext^{fea}$ is reduced by using the proposed A-PCA technique. The A-PCA efficiently reduces the proportionality of the features without information loss. Therefore, the function of the proposed A-PCA is already derived in section "Dimensionality reduction". Lastly, the dimensionality-reduced features are described as,

| Dataset | Feature name | Description | Justification |
|---|---|---|---|
| DPD | Gender | Male/female | Used for gender-based risk differences in diabetes prediction |
| | Age | Patient's age in years | Older individuals have a higher risk of diabetes, so it is considered for the analysis |
| | Hypertension | Presence of high blood pressure (yes/no) | Hypertension is a common comorbidity with diabetes |
| | Heart disease | History of heart disease (yes/no) | Cardiovascular complications are common in diabetic patients |
| | Smoking history | Never, former, current smoker | Smoking affects insulin resistance and diabetes risk |
| | BMI | Body mass index (BMI) (kg/m$^2$) | Shows obesity, which is a significant risk factor for diabetes |
| | HbA1c level | Average blood sugar levels over 2–3 months (%) | The key indicator for diagnosing diabetes |
| | Blood glucose level | Fasting blood glucose level (mg/dL) | High blood glucose levels are a primary diagnostic measure for diabetes |
| | Class label | Diabetes or no-diabetes | Presence of diabetes output classification (diabetic/non-diabetic) |
| SPD | ID | Unique identifier for each patient | Required for patient tracking but not used in prediction |
| | Gender | Male/female | Stroke risk varies between genders due to hormonal and lifestyle differences |
| | Age | Patient's age in years | Stroke occurs mostly in aged people |
| | Hypertension | Presence of high blood pressure (yes/no) | Hypertension is the leading cause of stroke |
| | Heart disease | History of heart disease (yes/no) | Stroke risk increases significantly in patients with heart disease |
| | Work type | Type of employment (government, private, self-employed, etc.,) | Lifestyle in certain jobs increases stroke risk |
| | Marital status | Married or single | Social and psychological factors impact the stroke risk |
| | Residence type | Urban/rural | Lifestyle and access to healthcare changes between urban and rural areas |
| | Average glucose level | Patient's average blood glucose level (mg/dL) | High glucose levels increase the stroke risk |
| | BMI | Body mass index (BMI) (kg/m$^2$) | Obesity is a major risk factor for stroke |
| | Smoking status | Never, former, current smoker | Smoking increases the risk of stroke |
| | Class label | Stroke or normal | Stroke analysis output (stroke/no-stroke) |

**Table 2**. Features description.

$$D_y = \sum_{y=1}^{Y} (D_1, D_y, \ldots D_Y) \tag{34}$$

Here, $Y$ is the number of dimensionality-reduced features $D_y$.

*Stroke classification*

In this, the $D_y$ is given as input to the proposed PO-RSNN, which effectively classifies the collected data as strokes and non-strokes. The proposed PO-RSNN process is already explained in Sect. 3.5.6. Hence, the classified outcome $(Cls_{out})$ is displayed as,

$$Cls_{out} \rightarrow \left\{stk, non^{stk}\right\} \tag{35}$$

Here, $stk$ and $non^{stk}$ depict the stroke and non-stroke class, correspondingly. Also, the diagnosed outcomes are encrypted and then securely stored in the BC. The research framework effectively provides communication support for disabled people and normal people in ME. In addition, the diabetes individuals are significantly classified by the proposed PO-RSNN, enhancing the model's reliability.

## Results and discussion

The objective of this section is to evaluate the significance of the research methodology. The proposed system is deployed on the working platform of PYTHON. The software requirement of the proposed work is given below,

The proposed system is developed using Python 3.10, which is a widely used programming language for DL applications. The PyCharm 2022.3.3 is utilized for coding, debugging, and executing the model efficiently on **a** Windows 10 (64-bit) platform. Several Python libraries play a crucial role in different aspects of the model. Here, Tkinter is used to create a user-friendly graphical interface, and Matplotlib is incorporated for visualizing datasets, model performance metrics, and the comparative analysis of different classification techniques. For DL capabilities, Keras and TensorFlow serve as the primary frameworks, improving the optimization of the predictive model. Next, the NumPy provides support for numerical computations, and the CSV library is used to read and process structured datasets. Then, the imblearn is used to address the class imbalance, and the Scikit-learn (sklearn) is utilized for model selection and training. Further, cryptography ensures the data security and integrity of the proposed model.

### Dataset description

The proposed work is evaluated by using two datasets like diabetes prediction dataset (DPD) and the stroke prediction dataset (SPD). The datasets are collected from the publically available resources that are mentioned under the reference section. The medical and demographic information of the subjects are included in the DPD

| Classes | Training | Testing | Total |
|---|---|---|---|
| Stroke | 3888 | 973 | 4861 |
| Non-Stroke | 3888 | 973 | 4861 |
| Total class | 7776 | 1946 | 9722 |

**Table 3**. Samples of the DPD.

| Classes | Training | Testing | Total |
|---|---|---|---|
| Diabetes | 73,200 | 18,300 | 91,500 |
| Non-diabetes | 73,200 | 18,300 | 91,500 |
| Total class | 1,46,400 | 36,600 | 1,83,000 |

**Table 4**. Samples of the SPD.

dataset that is utilized for predicting diabetes individuals. Moreover, features like gender, age, BMI, heart disease, hypertension, smoking history, HbA1c level, and blood glucose level are present in the DPD dataset. Likewise, the SPD is established to predict whether a patient is diagnosed with stroke or not centered on features like age, gender, various diseases, and smoking status. Furthermore, to enhance the classification rate, the classes of the dataset are balanced. Table 3 and Table 4 display the characteristics of the DPD and SPD, respectively,

From the whole data, 80% of the data is utilized to train the model and the remaining data is allocated for testing purposes. The hardware requirements of the proposed system are given as follows,

- Processor: Intel i5/ core i7
- CPU speed: 3.20 GHz
- OS: Windows 10
- System type: 64 bit
- RAM: 8 GB

## Performance analysis

Here, to reveal the supremacy of the system, the performance validation and comparative analysis are done for the proposed technique.

### Performance evaluation of classification (diabetes dataset)

Here, the performance of the proposed PO-RSNN is analogized with prevailing approaches, such as RNN, Deep Belief Network (DBN), Restricted Boltzmann Machine (RBM), DNN, ANN, and SVM.

The performance validation of the proposed PO-RSNN and existing approaches is illustrated in Fig. 3. The standard activation functions cause vanishing gradient issues, leading to poor generalization. Thus, the SAF is utilized in the proposed model, which ensures a consistent learning process that prevents excessive weight updates and reduces the overfitting issue. The overreliance on specific neurons is prevented, and this improves the model generalization. In addition, the weight parameter is optimized by utilizing the P-DBO; thus, the classification results are upgraded regarding the above salient features. Accuracy, precision, recall, training time, False Positive Rate (FPR), and False Negative Rate (FNR) attained by the proposed PO-RSNN are 98.97%, 98.96%, 98.97%, 58439 ms, 0.02, and 0.01, respectively. Similarly, the prevailing approaches achieved an average accuracy, precision, recall, training time, FPR, and FNR of 93.27%, 92.49%, 92.77%, 82566 ms, 0.209, and 0.284, respectively. The low values of FPR and FNR in the proposed model show that the model scales well while preserving the low misclassification rates. This ensures reliability and real-world applicability in DP. As the overfitting issue was rectified and the weight parameters were optimized, the result metrics for the proposed model were better than the traditional models. Thus, the proposed technique had a higher superiority in DP when compared to the existing methods.

### Performance evaluation of classification (stroke dataset)

Moreover, the performance of the proposed PO-RSNN is assessed by analogizing it with conventional classifiers based on accuracy, f-measure, sensitivity, specificity, True Positive Rate (TPR), and True Negative Rate (TNR).

Figure 4 represents the accuracy, f-measure, sensitivity, specificity, TPR, and TNR evaluation of the proposed PO-RSNN and existing techniques. To optimize the weight parameter, the proposed P-DPO is utilized, which elevates the convergence rate. The proposed PO-RSNN attained 98.89% accuracy, 98.97% f-measure, 98.99% sensitivity, 98.93% specificity, 98.52% TPR, and 98.48% TNR. But, the traditional works achieved poor performance. The higher TPR and TNR values attained by the proposed work prove that the likelihood of false alarms has been reduced, thus stating the reduction of misclassification. This proves trust in the analysis of the medical data in the proposed system. Thus, the proposed system obtained more accurate results in stroke prediction.

The model performance was evaluated using accuracy, precision, recall, and F-Measure. Each metric was chosen to provide a comprehensive assessment of classification effectiveness. Accuracy was used as a performance measure, and precision was considered as crucial to minimize false positives, particularly for

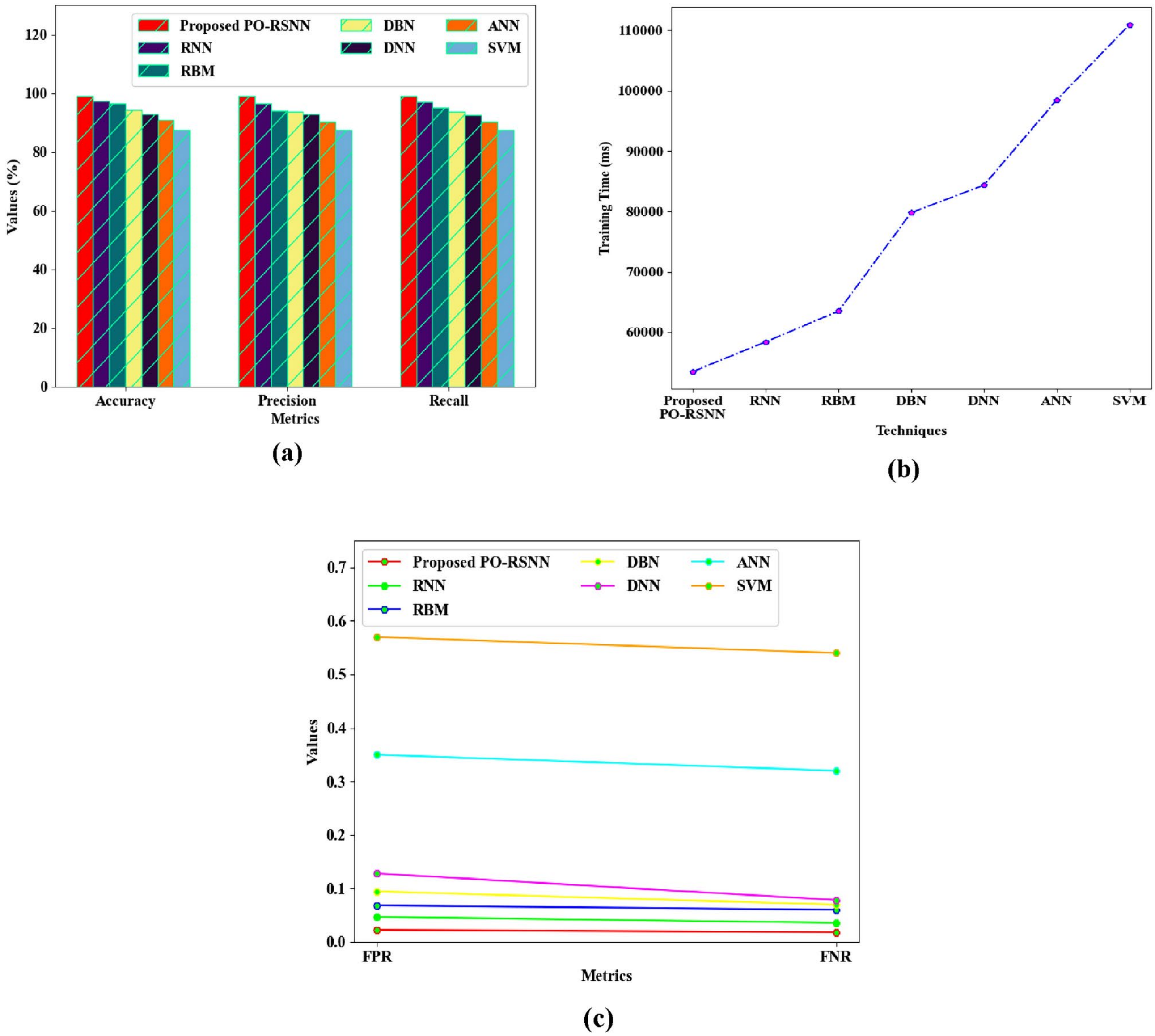



**Fig. 3**. Performance analysis of the proposed PO-RSNN regarding (**a**) accuracy, precision, and recall, (**b**) training time, and (**c**) FPR and FNR.

the medical diagnosis, where a false diagnosis could lead to unnecessary interventions. The recall metric was prioritized to ensure that actual positive cases of the diabetic class were crucially identified, thus reducing the risk of undetected conditions. The F-Measure score was selected as a balanced metric that harmonized precision and recall, thus making it suitable for scenarios with class imbalances. These metrics collectively ensured a robust evaluation framework, balancing trade-offs between sensitivity and specificity. These metrics were essential in DP, where both false positives and false negatives carried significant consequences.

## Comparative assessment of data security

Likewise, the proposed MJCC's performance is analogized with traditional techniques, namely ECC, Rivest, Shamir, Adleman (RSA), Diffie Helman (DH), ElGamal, Data Encryption Standard (DES), and Advanced Encryption Standard (AES).

Figure 5 exhibits the encryption time and decryption time analysis of the proposed MJCC. To increase the security level of the system, the proposed MJCC employs a Montwisted-Jaco curve. To encrypt and decrypt the data, the proposed MJCC takes 1132 ms and 1140 ms, respectively. But, to perform encryption and decryption, the prevailing approaches consume maximum time. As per the analysis outcomes, the proposed system had low time complexity and high privacy.

In Table 5, the Security Level (SL) of the proposed MJCC is validated by comparing it with conventional techniques. The proposed MJCC attained 98.92% SL, and the prevailing works obtained an average SL of 91.38%. According to Table 3, the proposed framework has performed well in secure CD monitoring.

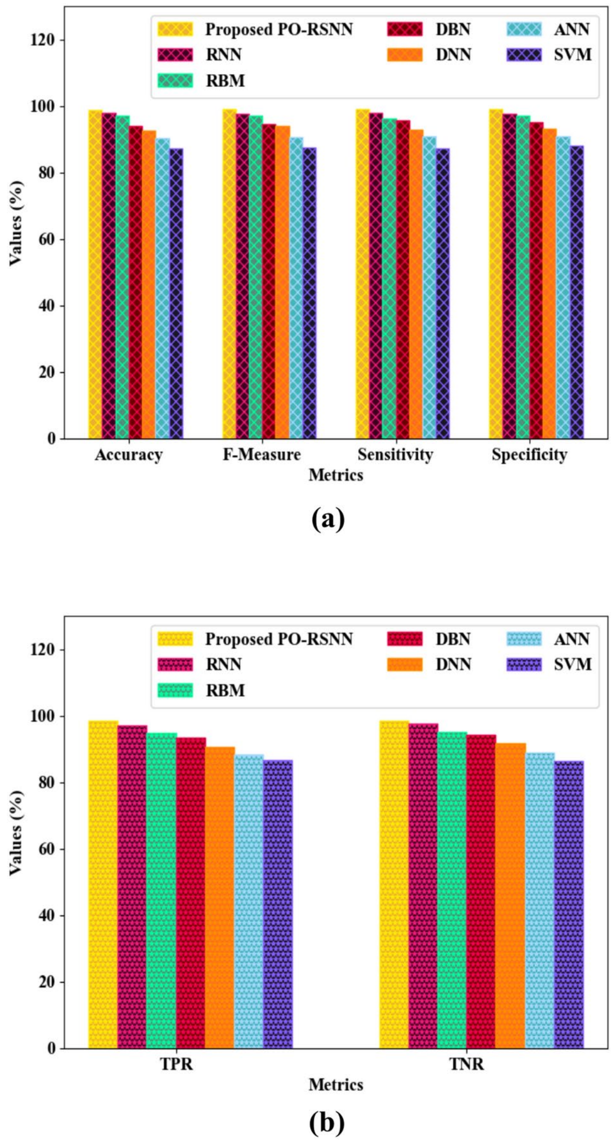



**Fig. 4**. Performance analysis regarding (**a**) accuracy, f-measure, sensitivity, specificity, and (**b**) TPR and TNR.

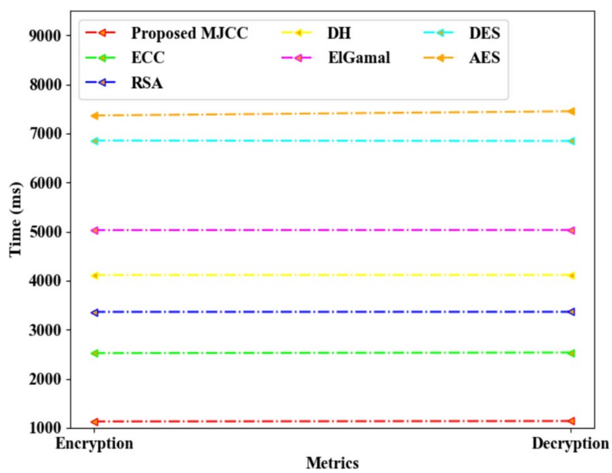



**Fig. 5**. Encryption and decryption time analysis.

| Techniques | Security level (%) |
|---|---|
| Proposed MJCC | 98.92 |
| ECC | 95.38 |
| RSA | 92.25 |
| DH | 89.65 |
| ElGamal | 88.27 |
| DES | 86.45 |
| AES | 84.21 |

**Table 5**. Security level analysis.

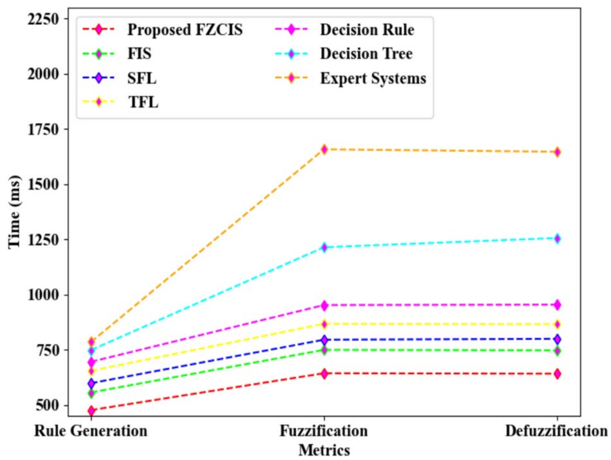



**Fig. 6**. Performance analysis of the proposed FZCIS.

| Methods | Prediction rate (%) |
|---|---|
| Proposed FZCIS | 98.78 |
| FIS | 96.14 |
| SFL | 93.45 |
| TFL | 91.02 |
| Decision rule | 89.87 |
| Decision tree | 87.25 |
| Expert systems | 85.47 |

**Table 6**. Prediction rate evaluation.

## Performance analysis of severity level estimation

Here, the performance analysis of the proposed FZCIS is done to reveal the model's consistency. Furthermore, the existing approaches, namely FIS, Sigmoid Fuzzy Logic (SFL), Trapezoidal Fuzzy Logic (TFL), Decision Rule (DR), Decision Tree, and Expert Systems are analogized with the proposed model.

The performance evaluation of the proposed FZCIS and existing techniques based on Rule Generation Time (RGT), Fuzzification Time (FT), and Defuzzification Time (DT) is displayed in Fig. 6. To perform defuzzification, the proposed framework uses z-log-clipping normalization, which increases the model's reliability. Z-log-clipping normalization indicates that the proposed approach attained an RGT, FT, and DT of 476 ms, 643 ms, and 641 ms, correspondingly. Likewise, the conventional methods acquired an average RGT, FT, and DT of 625 ms, 840 ms, and 841 ms, respectively. Therefore, the experimental results stated that when compared to prevailing algorithms, the proposed model achieved higher efficiency in severity analysis.

Table 6 demonstrates the Prediction Rate (PR) analysis of the proposed FZCIS and traditional approaches. The proposed FZCIS attained 98.78% PR. Similarly, the prevailing approaches attained an average PR of 90.53%. Thus, when compared to other prevailing models, the proposed work attained impressive outcomes.

## Performance evaluation of optimization

Additionally, the performance of the proposed P-DBO is assessed by comparing it with prevailing systems, such as DBO, African Vultures Optimization Algorithm (AVOA), Bees Algorithm (BA), Egret Swarm Optimization

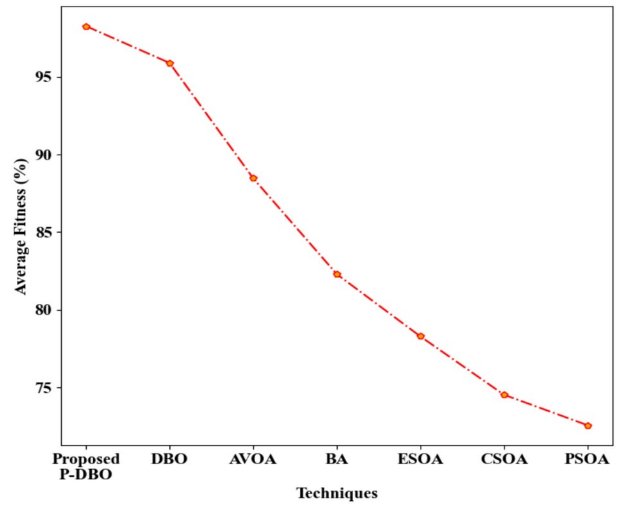



**Fig. 7**. Average fitness analysis.

| Works | Techniques | Accuracy (%) | F-measure (%) | RMSE |
|---|---|---|---|---|
| Proposed system | PO-RSNN and FZCIS | 98.97 | 98.96 | 1.78 |
| (Ihalapathirana et al. 2023)[37] | Hybrid-ML | 69 | 76 | – |
| (Malibari 2023)[38] | EO-LWAMCNet | 94 | 96 | – |
| (Nasser et al. 2021)[39] | RNN-RBM | – | – | 15.58 |
| (M. Chen et al. 2021)[40] | Hybrid-ML | 80 | 79 | – |
| (Thamotharan et al. 2023)[41] | LB-MPC | – | – | 5.01 |

**Table 7**. Comparative evaluation.

Algorithm (ESOA), Crow Search Optimization Algorithm (CSOA), and Particle Swarm Optimization Algorithm (PSOA).

In Fig. 7, the Average Fitness Value (AFV) of the proposed P-DBO and traditional techniques are analyzed. To perform the foraging behavior of the DB, the proposed work uses probabilistic distribution, thus enhancing the model's performance. The proposed P-DBO achieved an AFV of 98.26. Similarly, the traditional approaches acquired an average AFV of 82. Hence, the performance analysis displayed that the proposed work had high dominance in ME-centric CD prediction.

The proposed P-DBO employed computationally intensive methods, such as Dung Beetle Optimization (DBO) for weight adjustment and A-PCA for dimensionality reduction. While these techniques significantly enhance accuracy and efficiency, they may pose challenges for deployment in resource-constrained environments, such as edge devices and low-power computing platforms. Thus, the dimensionality reduction and the probability distribution in the proposed P-DBO were conducted with limited RAM and processing power.

## Comparative analysis of the proposed model

Based on quality metrics, the performance of the research framework is validated with related works.

In Table 7, the proposed work is compared with other associated models centered on their techniques and findings. The P-DBO-based weight optimization and SAF are utilized in the proposed PO-RSNN, which enhances the model's dependability. Then, the severity level of the diabetes patients is effectively estimated by the proposed FZCIS. In general, to predict chronic disorders, the prevailing works like learning-based model predictive control (LB-MPC), hybrid-ML, and equilibrium optimizer-based lightweight automatic modulation classification networks (EO-LWAMCNet) are utilized. Accuracy, f-measure, and root mean squared error (RMSE) attained by the proposed model are 98.97%, 98.96%, and 1.78, respectively. However, the conventional models have limited performance. Thus, the proposed method is proved as a less error-prone model.

Table 8 describes the comparison of the DPD used in the proposed system and another dataset named as Pima Indians Diabetes Database. The DPD attained an accuracy of 98.97%, precision of 98.96%, recall of 98.97%, and training time of 53489 ms. However, by applying the Pima Indians Diabetes Database, the accuracy, precision, recall, and training time of the other dataset were 97.24%, 97.17%, 96.99%, and 78347 ms, respectively. Thus, the analysis of the DP using the proposed work's data was better than the other one.

| Metrics | DPD | Pima Indians diabetes database |
|---|---|---|
| Accuracy (%) | 98.97 | 97.24 |
| Precision (%) | 98.96 | 97.17 |
| Recall (%) | 98.97 | 96.99 |
| Training time (ms) | 53,489 | 78,347 |

**Table 8**. Comparison with other dataset.

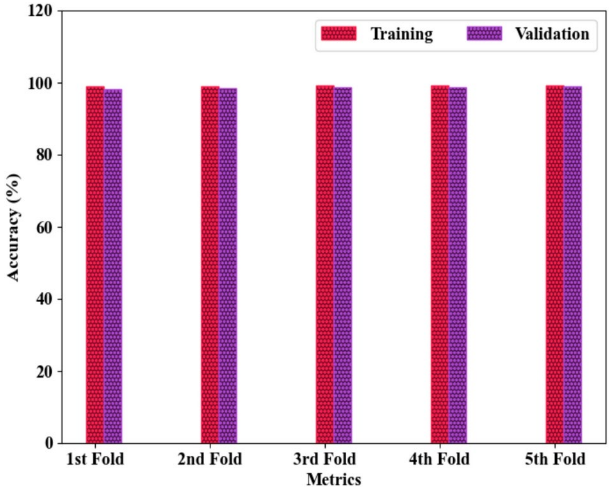



**Fig. 8**. Cross-validation.

## Generalizability of the model

The proposed work is designed to ensure high generalizability across the diverse differently abled persons like deaf and dumb individuals. A well-generalized model performs well on new and unseen data, and the validation of multiple data is given in Fig. 8.

As given in Fig. 8, the tenfold cross-validation strategy is applied. Here, the dataset was repeatedly divided into different training and testing splits rather than relying on a single training and testing partition. The model exhibited high stability across different folds, with a training accuracy ranging between 99.04% and 99.24% and validation accuracy ranging between 98.21% and 98.86% for the first five folds. The average validation accuracy of 98.58% proved that the model effectively learned the patterns from the training data and generalized the well-to-unseen samples. This strong cross-validation performance proves that the proposed PO-RSNN does not suffer from overfitting and is reliably used for real-world CD prediction.

Moreover, to address the issue of data imbalance, the SMOTE technique was applied in the proposed system. By generating synthetic instances of the minority class, the SMOTE analysis ensured a more balanced distribution of diabetic and non-diabetic cases. This enhanced the robustness of the DL model and also minimized the classification bias. Thus, the data balancing approach increased the generalizability of the model and prevented it from misclassification.

## Interpretability of recommendations and decision-making

In AI-driven healthcare applications, interpretability plays a vital role in ensuring that the model prediction is understandable and usable for making treatment plans. The proposed system not only predicts the DP and stroke analysis but also provides meaningful recommendations based on the patient's risk factor. This is attained through the combination of DL-based decision-making and fuzzy inference for severity estimation. The important features are selected and utilized by the DL architecture. Thus, based on the estimation, personalized recommendations, such as immediate medical consultation for high-risk patients, diet and exercise recommendations for moderate-risk patients, and routine check-ups for low-risk patients can be recommended effectively.

## Computational overhead and complexity analysis

In the DL models, computational efficiency plays a pivotal role in real-time applications, such as healthcare systems. The computational overhead and the complexity attained by the proposed PO-RSNN are given below,

As given in Fig. 9, the execution time regarding the computational overhead required for the processing of the patient's data in the proposed PO-RSNN was compared with the prevailing models, such as RNN, RBM, DBN, DNN, ANN, and SVM. The PO-RSNN model attained an execution time of 3,856 ms. However, the prevailing RNN, RBM, DBN, DNN, ANN, and SVM attained execution times of 7451 ms, 11254 ms, 14896 ms,

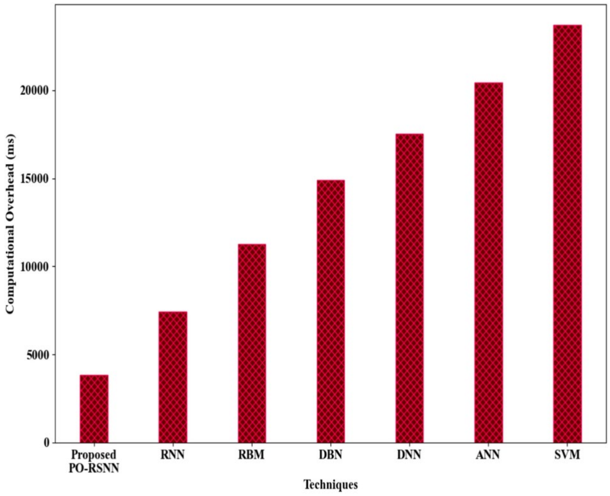



**Fig. 9**. Computational overhead analysis.

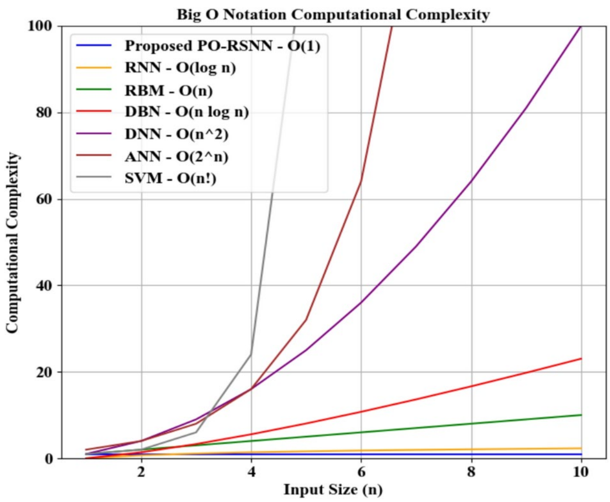



**Fig. 10**. Complexity analysis of PO-RSNN.

17524 ms, 20457 ms, and 23695 ms, respectively. The reduction in execution time proved that the PO-RSNN was computationally efficient for real-time healthcare applications.

Figure 10 displays the training time complexity of the proposed PO-RSNN and the traditional techniques using Big O notation. Normally, the Big O notation is used to define the upper bound running time of the required proposed method, and this gives a theoretical estimation of the performance of the proposed technique regarding the input size. Here, the proposed PO-RSNN had O(1) training time complexity because the proposed model overcame the vanishing gradient and overfitting issues. But, the prevailing RNN, RBM, DBN, DNN, ANN, and SVM techniques attained a time complexity of O(log n), O(n), O(n log n), O(n^2), O(2^n), and O(n!), respectively. Thus, the existing models attained higher time complexity than the proposed technique. This showed that the proposed technique had a lower time complexity than the existing classifiers.

## Real-wold security threat analysis

The security level of the proposed MJCC regarding the Eavesdropping Attack, Man-in-the-Middle (MitM) Attack, Replay Attack, Smurf Attack, Slowloris Attack, TCP Reset Attack, and ICMP Redirect Attack is given in Fig. 11.

The security level of the MJCC encryption technique in the proposed system demonstrates strong encryption protocols against eavesdropping attacks (99.0235%), MitM attacks (99.1245%), replay attacks (98.8745%) Smurf attacks (98.9635%), Slowloris attack (98.8562%), TCP Reset attacks (99.0247%), and ICMP attacks (98.4512%). This guaranteed the reduction of risks associated with malicious rerouting in the network traffic during patient's medical data transmission.

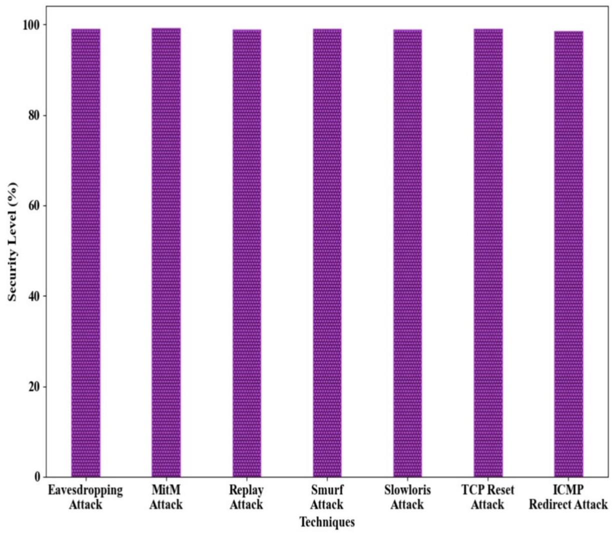



**Fig. 11**. Threat analysis for MJCC.

## Feature distribution analysis

The feature distribution of the proposed work with respect to the features that are processed before and after pre-processing for the two datasets, DPD and SPD, is given below,

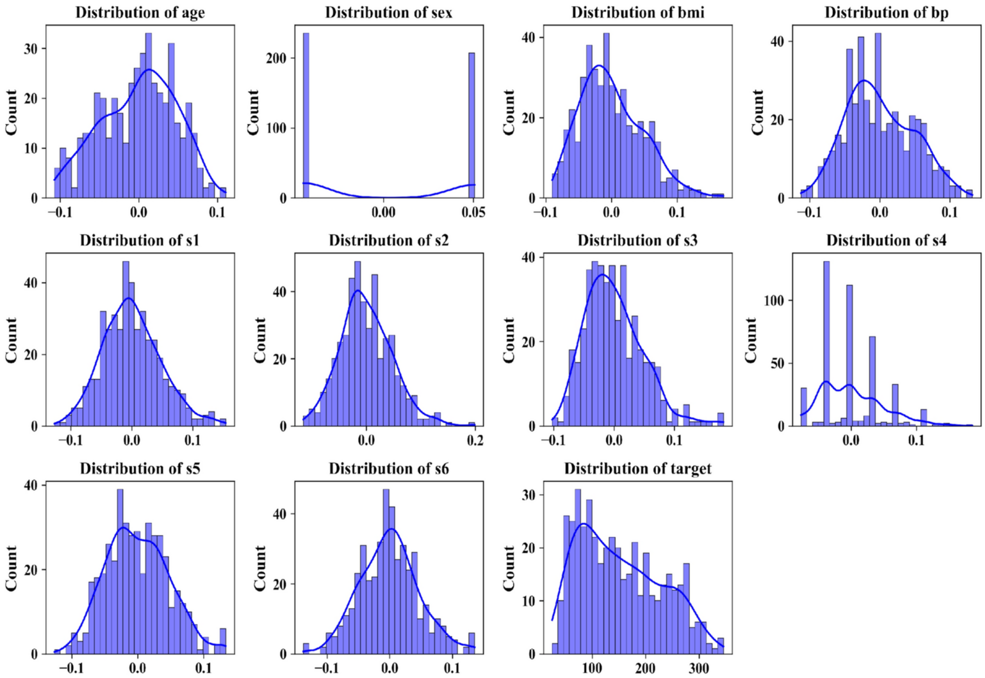
Figure 12a represents the feature distribution of the DPD dataset before preprocessing, showing significant variability and potential outliers. After applying pre-processing, Fig. 12b displays a more uniform and well-structured distribution. This proves that the preprocessing effectively enhanced data quality. Similarly, Fig. 13a highlights raw, skewed features, while Fig. 13b demonstrates a refined and balanced distribution for SPD. Thus, it is confirmed that the preprocessing steps in the proposed system improve data consistency and enhance the model interpretability.

## Conclusion

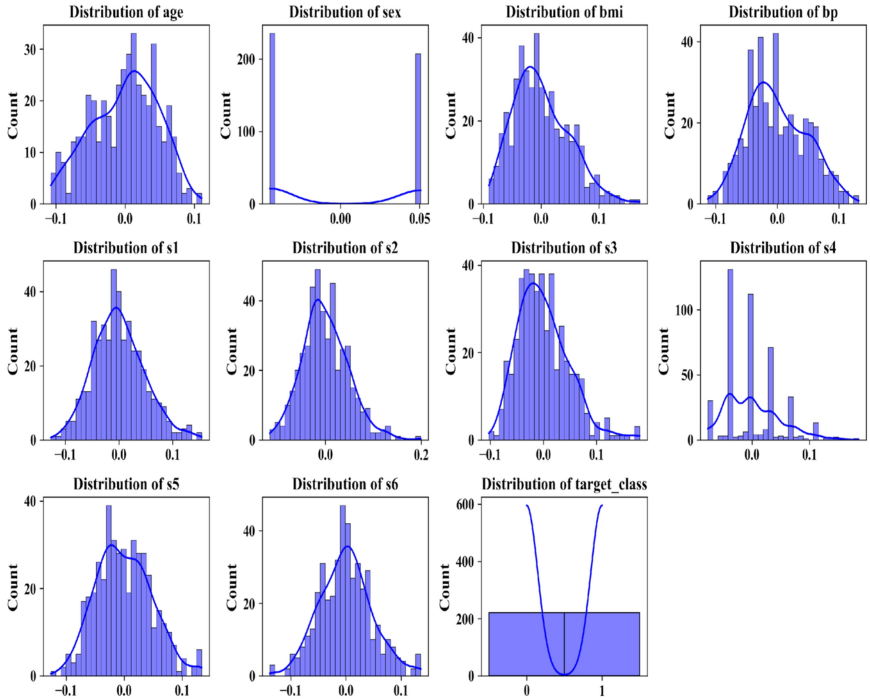
This paper proposed a novel PO-RSNN and FZCIS-based diabetes and stroke analysis framework in ME. For disabled patients, the proposed work provided significant communication support. Moreover, the diabetes and severity level were effectively predicted by the proposed PO-RSNN and FZCIS, correspondingly. Also, the research framework was implemented by using datasets like DPD and SPD. Hence, the performance analysis showed that for the DPD and SPD, the proposed PO-RSNN attained accuracies of 98.97% and 98.89%, respectively. In addition, the proposed FZCIS acquired 98.78% PR and displayed higher efficiency. Likewise, the proposed MJCC achieved 98.92% SL, which exhibited better security. Also, for all the quality metrics, the proposed system had high supremacy. Hence, the experimental analysis proved that the proposed framework outperformed the prevailing works in CD monitoring. The research approach rendered detailed insight into the communication strategies for disabled persons in ME. But, the proposed system only focused on communication support for disabled persons in ME and limited CD prediction.

## Limitations and future scope

Although the proposed work effectively predicted diabetes and analyzed stroke in a patient, other CDs, such as hypertension, cancer, and heart disease, were not considered. Thus, its applicability in a broader healthcare context was limited. So, in the future, the proposed work will be enhanced by focusing on other chronic illnesses, such as hypertension, cancer, and heart diseases utilizing more advanced DL techniques.

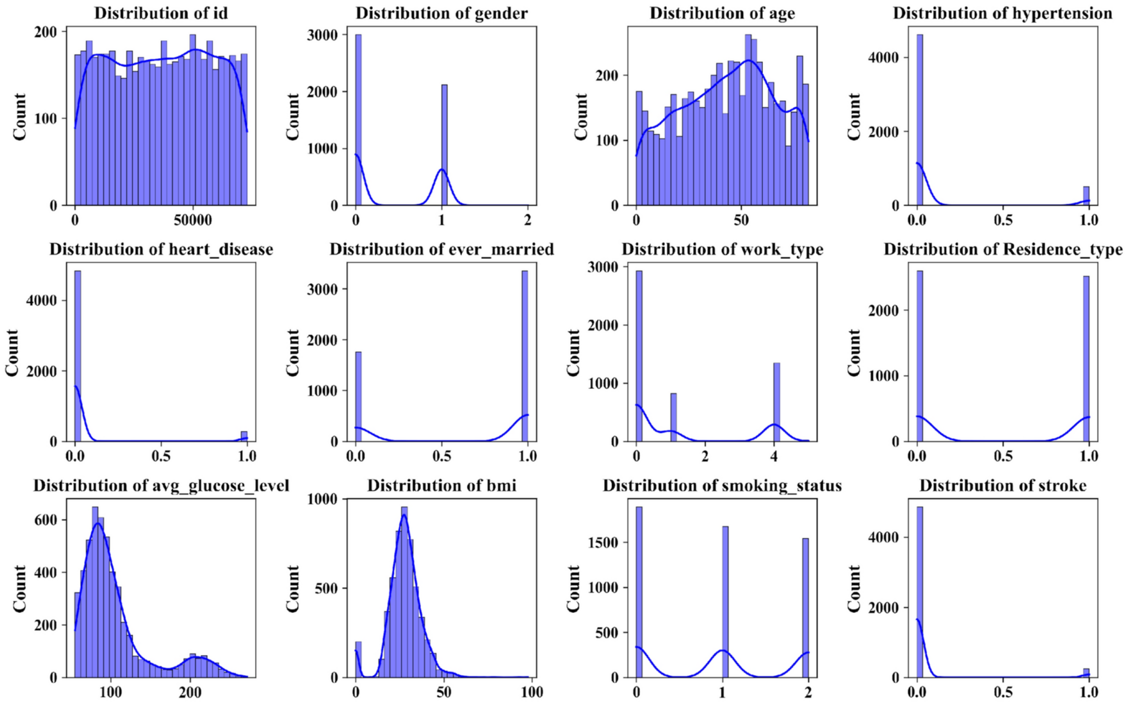



(a)

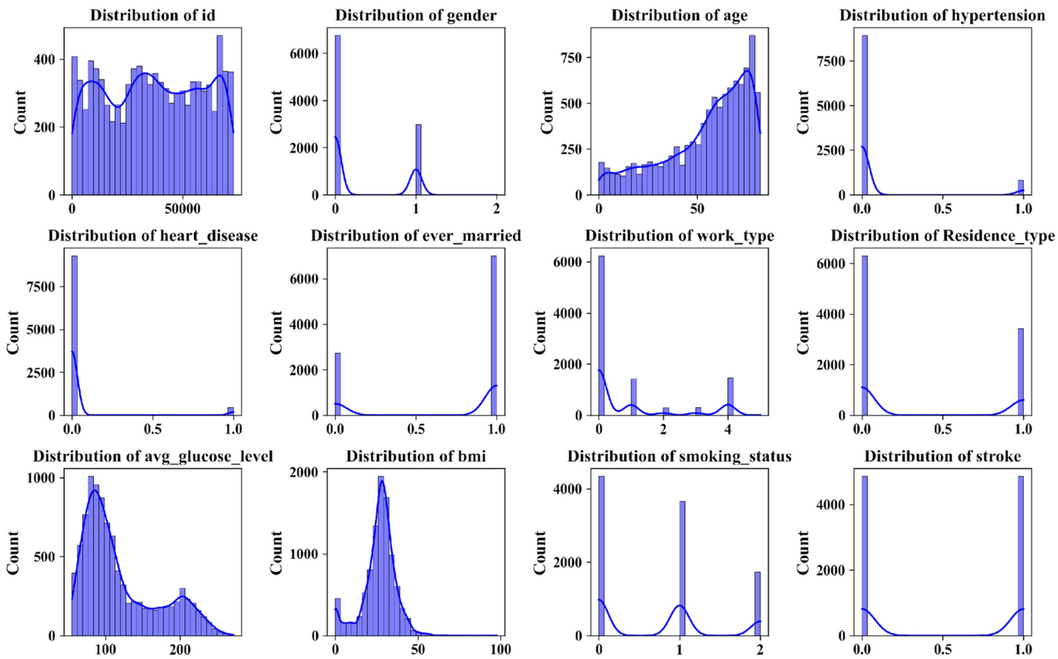



(b)

**Fig. 12**. Feature distribution of the DPD (**a**) Before pre-processing and (**b**) after pre-processing.




(a)




(b)

**Fig. 13**. Feature distribution of the SPD (**a**) Before pre-processing and (**b**) after pre-processing.

## Data availability

The data used to support the findings of this study are available from the corresponding author upon request. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset, https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset.

Received: 22 December 2024; Accepted: 28 March 2025
Published online: 04 April 2025

## References


1. Kumar, G. S. et al. SRADHO: Statistical reduction approach with deep hyper optimization for disease classification using artificial intelligence. *Sci. Rep.* **15**(1), 1245. https://doi.org/10.1038/s41598-024-82838-1 (2025).
2. Wu, J., Chang, L. & Yu, G. Effective data decision-making and transmission system based on mobile health for chronic disease management in the elderly. *IEEE Syst. J.* **15**(4), 5537–5548. https://doi.org/10.1109/JSYST.2020.3024816 (2021).
3. Yu, G. et al. Improving chronic disease management for children with knowledge graphs and artificial intelligence. *Expert Syst. Appl.* **201**, 1–12. https://doi.org/10.1016/j.eswa.2022.117026 (2022).
4. Wang, Y. C., Chen, T. C. T. & Chiu, M. C. A systematic approach to enhance the explainability of artificial intelligence in healthcare with application to diagnosis of diabetes. *Healthcare Anal.* **3**, 1–10. https://doi.org/10.1016/j.health.2023.100183 (2023).
5. Chaki, J., Thillai Ganesh, S., Cidham, S. K. & Ananda Theertan, S. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review. *J. King Saud Univ. Comput. Inf. Sci.* **34**(6), 3204–3225. https://doi.org/10.1016/j.jksuci.2020.06.013 (2022).
6. Kumar, A., Kumar, M., Bhardwaj, V. P., Kumar, S. & Selvarajan, S. A novel skin cancer detection model using modified finch deep CNN classifier model. *Sci. Rep.* **14**(1), 11235. https://doi.org/10.1038/s41598-024-60954-2 (2024).
7. Boukhennoufa, I., Zhai, X., Utti, V., Jackson, J. & McDonald-Maier, K. D. Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review. *Biomed. Signal Process. Control* **71**, 1–16. https://doi.org/10.1016/j.bspc.2021.103197 (2022).
8. Guan, Z. et al. Artificial intelligence in diabetes management: Advancements, opportunities, and challenges. *Cell Rep. Med.* **4**(10), 1–20. https://doi.org/10.1016/j.xcrm.2023.101213 (2023).
9. Singh, S. et al. Efficient pneumonia detection using vision transformers on chest X-rays. *Sci. Rep.* **14**(1), 2487. https://doi.org/10.1038/s41598-024-52703-2 (2024).
10. Chen, X. et al. Management for stroke intelligent early warning empowered by big data. *Comput. Electr. Eng.* **106**, 1–15. https://doi.org/10.1016/j.compeleceng.2023.108602 (2023).
11. Shamanna, P. et al. Reducing HbA1c in type 2 diabetes using digital twin technology-enabled precision nutrition: A retrospective analysis. *Diabetes Ther.* **11**(11), 2703–2714. https://doi.org/10.1007/s13300-020-00931-w (2020).
12. Hennebelle, A., Materwala, H. & Ismail, L. HealthEdge: A machine learning-based smart healthcare framework for prediction of type 2 diabetes in an integrated IoT, edge, and cloud computing system. *Procedia Comput. Sci.* **220**, 331–338. https://doi.org/10.1016/j.procs.2023.03.043 (2023).
13. Phan, H. L., Le, T. H., Lim, J. M., Hwang, C. H. & Koo, K. I. Effectiveness of Augmented Reality in Stroke Rehabilitation: A Meta-Analysis. *Appl. Sci. (Switzerland)* **12**(4), 1–17. https://doi.org/10.3390/app12041848 (2022).
14. Huang, M. et al. Design of a hybrid brain-computer interface and virtual reality system for post-stroke rehabilitation. *IFAC-PapersOnLine* **53**(2), 16010–16015. https://doi.org/10.1016/j.ifacol.2020.12.399 (2020).
15. Son, T. W., Ramli, D. A. & Aziz, A. A. Wearable heat stroke detection system in IoT-based environment. *Procedia Comput. Sci.* **192**, 3686–3695. https://doi.org/10.1016/j.procs.2021.09.142 (2021).
16. Shynu, P. G., Menon, V. G., Kumar, R. L., Kadry, S. & Nam, Y. Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing. *IEEE Access* **9**, 45706–45720. https://doi.org/10.1109/ACCESS.2021.3065440 (2021).
17. Kantipudi, M. P., Kumar, N. P., Aluvalu, R., Selvarajan, S. & Kotecha, K. An improved GBSO-TAENN-based EEG signal classification model for epileptic seizure detection. *Sci. Rep.* **14**(1), 843. https://doi.org/10.1038/s41598-024-51337-8 (2024).
18. Murala, D. K., Panda, S. K. & Dash, S. P. MedMetaverse: Medical care of chronic disease patients and managing data using artificial intelligence, blockchain, and wearable devices state-of-the-art methodology. *IEEE Access* **11**, 138954–138985. https://doi.org/10.1109/ACCESS.2023.3340791 (2023).
19. Ahmed, A., Aziz, S., Qidwai, U., Abd-Alrazaq, A. & Sheikh, J. Performance of artificial intelligence models in estimating blood glucose level among diabetic patients using non-invasive wearable device data. *Comput. Methods Progr. Biomed. Update* **3**, 1–7. https://doi.org/10.1016/j.cmpbup.2023.100094 (2023).
20. Asl, Z. R., Rezaee, K., Ansari, M., Zare, F. & Roknabadi, M. H. A. A review of biopolymer-based hydrogels and IoT integration for enhanced diabetes diagnosis, management, and treatment. *Int. J. Biol. Macromol.* https://doi.org/10.1016/j.ijbiomac.2024.135988 (2024).
21. Khalili Pour, E. et al. Automated machine learning–based classification of proliferative and non-proliferative diabetic retinopathy using optical coherence tomography angiography vascular density maps. *Graefe's Arch. Clin. Exp. Ophthalmol.* **261**(2), 391–399. https://doi.org/10.1007/s00417-022-05818-z (2023).
22. Annamalai, N. Diabetes mellitus prediction and severity level estimation using OWDANN algorithm. *Comput. Intell. Neurosci.* **2021**, 1–11. https://doi.org/10.1155/2021/5573179 (2021).
23. Hennebelle, A. et al. Secure and privacy-preserving automated machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction. *Comput. Struct. Biotechnol. J.* **23**, 212–233. https://doi.org/10.1016/j.csbj.2023.11.038 (2024).
24. Razfar, N., Kashef, R. & Mohammadi, F. An Artificial Intelligence model for smart post-stroke assessment using wearable sensors. *Decis. Anal. J.* **7**, 1–13. https://doi.org/10.1016/j.dajour.2023.100218 (2023).
25. Huang, R. et al. An artificial intelligence diabetes management architecture based on 5G. *Digital Commun. Netw.* https://doi.org/10.1016/j.dcan.2022.09.004 (2022).
26. Prabhakar, G., Chintala, V. R., Reddy, T. & Ruchitha, T. User-cloud-based ensemble framework for type-2 diabetes prediction with diet plan suggestion. *E-Prime Adv. Electr. Eng. Electron. Energy* **7**, 1–10. https://doi.org/10.1016/j.prime.2024.100423 (2024).
27. Mahesh, T. R. et al. Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease. *Comput. Intell. Neurosci.* https://doi.org/10.1155/2022/4451792 (2022).
28. Allen, A. et al. A digital twins machine learning model for forecasting disease progression in stroke patients. *Appl. Sci. (Switzerland)* **11**(12), 1–14. https://doi.org/10.3390/app11125576 (2021).
29. Lu, T. & Wang, Y. Prediction model construction for ischemic stroke recurrence with BP network and multivariate logistic regression and effect of individualized health education. *Comput. Math. Methods Med.* **2022**, 1–10. https://doi.org/10.1155/2022/4284566 (2022).
30. Elbagoury, B. M. et al. A hybrid stacked CNN and residual feedback GMDH-LSTM deep learning model for stroke prediction applied on mobile AI smart hospital platform. *Sensors* **23**(7), 1–35. https://doi.org/10.3390/s23073500 (2023).
31. Krishnamoorthi, R. et al. Retracted: A novel diabetes healthcare disease prediction framework using machine learning techniques. *J. Healthc. Eng.* https://doi.org/10.1155/2023/9872970 (2023).
32. Rastogi, R. & Bansal, M. Diabetes prediction model using data mining techniques. *Meas. Sens.* **25**, 100605. https://doi.org/10.1016/j.measen.2022.100605 (2023).
33. Zhou, H., Xin, Y. & Li, S. A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinform.* **24**(1), 224. https://doi.org/10.1186/s12859-023-05300-5 (2023).
34. Al Reshan, M. S. et al. An innovative ensemble deep learning clinical decision support system for diabetes prediction. *IEEE Access* https://doi.org/10.1109/ACCESS.2024.3436641 (2024).

35. Sai, M. J. et al. An ensemble of light gradient boosting machine and adaptive boosting for prediction of type-2 diabetes. *Int. J. Comput. Intell. Syst.* **16**(1), 14. https://doi.org/10.1007/s44196-023-00184-y (2023).
36. Chee, L. Z., Sivakumar, S., Lim, K. H. & Gopalai, A. A. Gait acceleration-based diabetes detection using hybrid deep learning. *Biomed. Signal Process. Control* **92**, 105998. https://doi.org/10.1016/j.bspc.2024.105998 (2024).
37. Ihalapathirana, A. et al. Explainable artificial intelligence to predict clinical outcomes in type 1 diabetes and relapsing-remitting multiple sclerosis adult patients. *Inform. Med. Unlock.* **42**, 1–15. https://doi.org/10.1016/j.imu.2023.101349 (2023).
38. Malibari, A. A. An efficient IoT-artificial intelligence-based disease prediction using lightweight CNN in healthcare system. *Meas. Sens.* **26**, 1–9. https://doi.org/10.1016/j.measen.2023.100695 (2023).
39. Nasser, A. R. et al. IoT and cloud computing in health-care: A new wearable device and cloud-based deep learning algorithm for monitoring of diabetes. *Electronics (Switzerland)* **10**(21), 1–12. https://doi.org/10.3390/electronics10212719 (2021).
40. Chen, M. et al. Blockchain-enabled healthcare system for detection of diabetes. *J. Inf. Secur. Appl.* **58**, 1–12. https://doi.org/10.1016/j.jisa.2021.102771 (2021).
41. Thamotharan, P. et al. Human Digital Twin for Personalized Elderly Type 2 Diabetes Management. *J. Clin. Med.* **12**(6), 1–36. https://doi.org/10.3390/jcm12062094 (2023).

## Author contributions

S.S, K.M, A.M and K.V wrote the main manuscript text, J.K and N.R prepared the figures K.M and V.B reviewed entire manuscript and prepared the table and technical methodology sections. All the authors are contributed equally for the manuscript preparation, review and submission.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.K. or V.B.