Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 31s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Deep Neural Network Approach for Early-Stage Diabetes Risk Prediction using Hybrid SMOTE-ENN and GAN with SHAP-Based Feature Explanations

Dr. Sheetalrani R Kawale[1], Pooja Kallappagol[2], Dr. Sharankumar Holyachi[3]

[1,2]Dept. of Computer Science,[3]Dept. of Community Medicine

[1,2]Karnataka State Akkamahadevi Women University, Vijayapura-586105, India, [3]Koppal Institute of Medical Sciences, Koppal, India

[1]Email ID: sheetalrani@kswu.ac.in, [2]Email ID: poojakallappagol30@gmail.com,

[3]Email ID: sharanholyachi2010@gmail.com

Cite this paper as: Dr. Sheetalrani R Kawale, Pooja Kallappagol, Dr. Sharankumar Holyachi, (2025) Deep Neural Network Approach for Early-Stage Diabetes Risk Prediction using Hybrid SMOTE-ENN and GAN with SHAP-Based Feature Explanations. *Journal of Neonatal Surgery*, 14 (31s), 861-874.

ABSTRACT

Type 2 diabetes mellitus (T2DM) is prevalent in India and remains a major global health concern. Timely recognition of type 2 diabetes is essential for successful management, especially in resource-constrained environments where access to standard laboratory testing might be restricted. This research introduces a deep learning framework aimed at predicting early-stage diabetes, emphasizing clarity and practical application in clinical environments through the use of non-invasive, symptom-based inputs. The Southern India Diabetes Dataset (SIDD) is a notable regional collection, comprising 1,680 ethically sourced patient records along with 17 clinical and demographic variables. A hybrid augmentation strategy was employed to address the class imbalance, incorporating SMOTE-ENN in aggregation with Generative Adversarial Networks (GANs). Furthermore, SHAP (SHapley Additive exPlanations) values were utilized to identify essential predictive features, enhancing the model interpretability. We employed and assessed two neural architectures: the Radial Basis Function Neural Network (RBFNN) and the Deep Neural Network (DNN). The proposed DNN model achieved a test accuracy of 98%, surpassing the performance of models trained on standard datasets such as PIMA. The proposed framework shows significant potential for application in essential healthcare settings, due to its incorporation of interpretable artificial intelligence, strong augmentation, and pertinent clinical data. This will facilitate prompt intervention and improve patient outcomes.

Keywords: *SMOTE-ENN · Generative Adversarial Networks · SHAP · Feature selection· Symptom-based screening · Regional health data · Interpretability · Neural networks*

1. INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder defined by a resistance to insulin and a developing dysfunction of pancreatic β-cells. It has become a significant non-communicable disease burden in India. This surge is fueled by a complex interplay of genetic predisposition and rapid lifestyle transitions associated with urbanization. The Indian Council of Medical Research–India Diabetes (ICMR–INDIAB) research, which was done with the Madras Diabetes Research Foundation (MDRF), says that over 74 million Indian people, or about 8.3% of the population, now have diabetes. Alarmingly, more than 57% of these individuals remain undiagnosed, underscoring the urgent need for improved nationwide screening, awareness, and preventive strategies [1]. Notably, the mean age of diabetes onset in India is approximately 42.5 years, significantly younger than the global average. This earlier onset increases the duration of disease exposure and consequently elevates the risk of chronic complications over a lifetime [2]. The urban-rural divide in prevalence is striking: urban regions report a diabetes prevalence of 16.4% compared to 8.9% in rural areas. This discrepancy reflects the influence of sedentary lifestyles, nutritional transitions, and differential access to healthcare services across India [3]. The distribution of diabetes across various states in India underscores pronounced regional disparities. Southern and coastal regions such as Goa (26.4%), Puducherry (26.3%), and Kerala (25.5%) report the highest prevalence, reflecting the impact of urbanization, dietary transitions, and sedentary lifestyles. Northern urban centers like Chandigarh (20.4%) and Delhi (17.8%) also exhibit elevated rates. Eastern and southern states including Tamil Nadu (14.4%) and West Bengal (13.7%) shows substantial prevalence, while northeastern states such as Sikkim (12.8%), Karnataka (14.8%), Tripura (9.4%), and Mizoram (8.5%) indicate moderate but concerning levels. In contrast, central and northern states Bihar (4.3%), Jharkhand (5.3%), Uttar Pradesh (4.8%), and Meghalaya (4.5%) report comparatively lower prevalence, although upward trends are emerging. These

figures highlight significant regional heterogeneity and emphasize the need for tailored public health strategies [4].The epidemic of diabetes is growing quickly over the world. Roughly 10.5% of the world's population, or roughly 537 million persons between the ages of 20 and 79, resided with diabetes in 2021. By 2030, this number is expected to grow to 643 million, and by 2045, it is expected to grow to 783 million [5]. In 2021, the prevalence in the South-East Asia (SEA) area was 8.8%, although India had a slightly higher rate of 9.6%. By 2045, projections show that SEA's rate would rise to 11.5% and India's rate will rise to 10.9%. This will be a major public health problem for the area. In addition to diagnosed diabetes, pre-diabetes, which is marked by elevated blood glucose levels that do not meet the criteria for diabetes diagnosis, represents a considerable concern. This high-risk metabolic state serves as a precursor to T2DM and presents a critical opportunity for intervention. Rooney et al. (2023) estimated that in 2021, 9.1% of the global adult population experienced impaired glucose tolerance (IGT), with projections indicating an increase to 10.0% by 2045. The global prevalence of impaired fasting glucose (IFG) stands at 5.8% and is projected to rise significantly. The data highlights the importance of early detection and proactive measures to prevent the advancement of disease [6]. Diabetes Mellitus (DM) is a widely recognized chronic condition characterized by sustained hyperglycemias, typically identified through biochemical assessments including fasting plasma glucose (FPG), glycated haemoglobin (HbA1c), and urine glucose tests [7]. Although these methods are considered the benchmark, they depend on clinical infrastructure and may face delays due to accessibility or cost challenges. To overcome these limitations, symptom-based screening enhanced by artificial intelligence has surfaced as a promising approach. Recent frameworks utilizing deep learning models have demonstrated significant accuracy in predicting diabetes risk by incorporating self-reported symptoms and demographic data, eliminating the necessity for immediate laboratory confirmation. In particular, a 2025 study proposed a context-aware dynamic ensemble model combining AdaBoost and support vector machines, trained on both global and localized datasets, achieving superior accuracy over traditional method [8]. This study seeks to build upon these innovations by introducing a deep learning-based diagnostic tool focused on symptomatology, offering an accessible and early-stage diabetes risk assessment system suitable for diverse healthcare environments. The structure of this study is organized to provide a coherent and methodical exploration of the proposed approach. It commences with a review of existing literature, offering foundational context and highlighting recent advancements in diabetes prediction using deep learning methodologies. Section 2 introduces the dataset utilized, including explanations on data preprocessing and primary data acquisition methods, and deep learning techniques applied. Section 3 will address experimental settings; the empirical result and the rigorous evaluation will be reported. Finally, the major results, implications of the findings, and lessons for future research are summarized in Section 4, and the paper is concluded with indications for further work.

2. RELATED WORK

Diabetes mellitus is a chronic disease requiring timely diagnosis and effective management. It is associated with serious complications such as cardiovascular disease, kidney failure, and nerve damage, making early detection essential. In recent years, machine learning, especially deep learning techniques like the Deep Belief Network (DBN), has proven to be a valuable method for the early prediction of diabetes.DBNs are a class of generative neural networks that use stacked Restricted Boltzmann Machines to learn hierarchical feature representations, making them particularly well-suited for complex biomedical data. Multiple studies have demonstrated the utility of DBNs in enhancing predictive accuracy for diabetes diagnosis across diverse populations and clinical settings [9-11]. This section synthesizes key research contributions on Deep learning models for diabetes diagnosis, with a focus on model accuracy and datasets used, to highlight the current state and effectiveness of these techniques in clinical informatics. Shahin et al. [9] employed a robust DBN model to forecast diabetes complications utilizing patient medical records. The model achieved a moderate accuracy of 81.25%, surpassing traditional machine learning algorithms, despite the specific dataset not being identified in the study. Prabhu and Selvabharathi [10] created a classifier based on DBN utilizing the Pima Indian Diabetes dataset. The model demonstrated superior performance compared to traditional classifiers such as Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, and SVM, achieving a recall of 1.0, precision of 0.6791, and an F1-measure of 0.808, thereby affirming its enhanced diagnostic capabilities. Panigrahy et al. [11] presented a hybrid model that integrates Deep Belief Networks (DBNs) with Tabu Search Optimization (TSO) for the purpose of hyperparameter tuning, referred to as TSO-DBN. The model effectively tackled class imbalance through the application of SMOTE, resulting in a prediction accuracy of 96.23%, an F1-score of 0.8749, and a Matthews Correlation Coefficient (MCC) of 0.8863. The dataset was sourced from clinical diabetes reports and was subjected to comprehensive preprocessing to enhance its quality. Lang et al. [12] created a DBN model that employs integrated algorithms and evaluated its performance against logistic regression and support vector machine (SVM). The dual-hidden-layer DBN model attained a maximum AUC of 81.62%, demonstrating superior classification performance compared to the alternatives. Nevertheless, accuracy values were not explicitly reported. Liu et al. [13] developed a diabetes management strategy based on a Deep Belief Network, utilizing the Pima Indians Diabetes dataset. The DBN model demonstrated an accuracy of 77.60%, surpassing the performance of backpropagation neural networks at 76.3% and support vector machines at 76.56%. The research further identified significant risk factors, including plasma glucose and BMI, by employing weight matrices across layers. Reddy et al. [14] employed the Pima Indian Diabetes dataset to forecast hospital readmission utilizing a Deep Belief Network (DBFN). The model attained an accuracy of 69.17%, surpassing the performance of Logistic Regression, Decision Tree, and Gradient Boosting. The results demonstrated notable performance

in specificity (66.44%) and NPV (70.32%), highlighting its effectiveness in clinical prediction tasks. Ma et al. [15] introduced a DBN model utilizing ReLU activation functions, which were trained through contrastive divergence and subsequently refined using backpropagation. The model attained an accuracy of 81% utilizing the Pima Indian Diabetes dataset. The focus was on integrating unsupervised and supervised learning to enhance classification performance. Olabanjo et al. [16] proposed an ensemble feature selection combined with unsupervised deep neural network (DBN) with a dataset prepared from the Sylhet Diabetes Hospital of Bangladesh. The effectiveness of the model in early identification of type 2 diabetes is indicated with an F1-score of 1.00, precision of 0.92, and recall of 1.00. Bala Manoj Kumar et al. (2020) [17] introduced a Deep Neural Network (DNN) classifier model that incorporates feature importance selection mechanisms, including Extra Trees and Random Forest algorithms. The methodology resulted in a notable enhancement in prediction accuracy, reaching 98.16% with an 80–20 train-test split on the PID dataset. Ashiquzzaman et al. (2017) [18] introduced a dropout-boosted Deep Learning Neural Network to address the challenge of over fitting in intricate neural models. The model included dropout layers between the fully connected layers, which facilitated multiple independent representations of the data. Their network attained an accuracy of 88.41% on the PID dataset directly at input, without the need for pre-processing and normalization. A comprehensive examination of previous research reveals several persistent shortcomings in the area of diabetes prediction. Initially, earlier studies have primarily relied on widely available benchmark datasets such as PIMA and UCI. While these datasets provide convenience, they do not accurately reflect real-world clinical environments or the demographic characteristics unique to various regions. Secondly, the feature selection methods and predictive models utilized in these studies were infrequently reviewed or validated by medical professionals, which raises concerns about their clinical relevance and diagnostic accuracy. Thirdly, a notable limitation exists in the inadequate comparison between model-generated predictions and actual clinical outcomes, including predicted probabilities and associated confidence scores. These factors are essential for evaluating the reliability, robustness, and practical applicability of such models in real-world healthcare settings.

3. PROPOSED METHOD

Our previous study used a dataset of 806 clinical database of Southern India Diabetes Dataset (SIDD). In this study, we expanded the dataset to 1,680 samples, thereby improving the prediction reliability and generalization. We present this work-in-progress article as introduction of a clinically motivated deep learning framework for the early predication of T2DM based on real world symptoms-based data, acquired under ethical clearance from Koppal Institute of Medical Sciences, Karnataka. The proposed approach overcomes limitations of prior works by integrating clinician validated clinical characteristics, interpreting the model outputs by explainable AI, and measuring the reliability of predictions using confidence scores. The enriched dataset contains 17 clinical and demographic features. To mitigate class imbalance, a hybrid data augmentation method which is a combination of SMOTE-ENN and GANs is employed. The SMOTE-ENN creates synthetic minority samples and eliminates noisy majority class cases at the same time. After that, GANs are utilized to promote diversity of minority samples. This two-stage enhancement produces a balanced dataset, which is compatible with deep learning. Then, SHAP (SHapley Additive exPlanations) values based on a Random Forest classifier are used to analyze feature importance. This approach provides interpretability and clinical relevance since only relevant and highly predictive features are considered. Such functions are additionally examined and verified by the medical authorities to determine their diagnostic power. Two neural models, namely a DNN implementing dropout and batch normalization as regularization solutions and an RBFNN employing Gaussian activation functions to model local data structure, are presented and trained. Performance of both models is assessed in terms of precision, recall, F1-score, accuracy and AUC-ROC, common metrics used for credibility analytics. Predictions and associated confidence scores are compared with clinical labels to confirm the robustness and clinical relevance of the model. The complete framework is shown in Fig. 1.

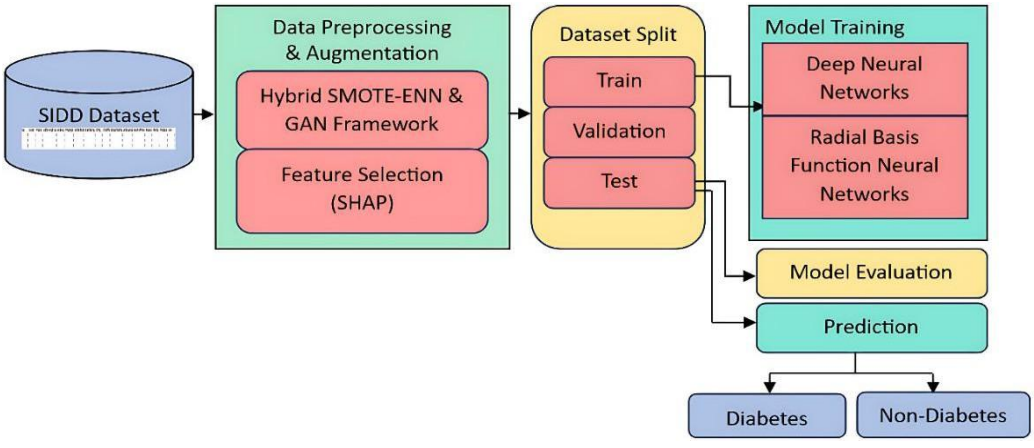


Fig 1: Proposed diabetes prediction framework integrating hybrid data augmentation, SHAP-based feature selection, and neural network classifiers

3.1 Primary Data Collection

A novel dataset, known as the Southern India Diabetes Dataset (SIDD), has been established to facilitate the development of a strong and clinically pertinent predictive framework. The data collection process was conducted at the Koppal Institute of Medical Sciences (KIMS) in Karnataka, India, adhering to the established institutional ethical clearance protocols to ensure compliance with patient confidentiality and data privacy standards. The dataset was assembled using a systematic symptom-reporting approach, in which each participant was asked about the presence or absence of particular clinical symptoms. Symptom responses were recorded as binary variables: "yes" indicating presence and "no" indicating absence, with diagnostic classification based on blood test-confirmed outcomes either "diabetic" or "non-diabetic." The final dataset consists of 1,680 patient records, each meticulously annotated with 17 clinical and demographic attributes. The features used in this study are the same as those employed in the early-stage diabetes risk prediction dataset [29], and they align with medically established indicators for early-stage Type 2 Diabetes Mellitus (T2DM). In contrast to standardized datasets like UCI, which primarily originate from patient surveys, the SIDD dataset is developed from clinical observations that have been validated by physicians. This guarantees an increased level of diagnostic reliability and improves the contextual significance of the dataset for the Southern Indian population.

3.2 Dataset Description

The dataset utilized in this study comprises 1,680 patient records, each annotated with 17 attributes that are clinically and demographically significant. The features encompass patient age, gender, and a thorough range of symptoms typically linked to Type 2 Diabetes Mellitus (T2DM). Among the total records, 66.31% are identified as diabetic cases according to clinical diagnosis, whereas the remaining 33.69% are categorized as non-diabetic instances. Table 1 presents a comprehensive summary of the dataset's attributes, encompassing symptom descriptors and demographic variables.

Table 1 Clinical symptoms and demographic characteristics included in the Southern India Diabetes Dataset (SIDD), used for early-stage Type 2 Diabetes Mellitus prediction

Sl. No	Features	Description
1	Age	Age of the individual (range: 14–85 years)
2	Gender	Biological sex (Male/Female)
3	Polyuria	Excessive urination (Yes/No)
4	Sudden Weight Loss	Unexplained weight loss (Yes/No)
5	Weakness	General body weakness (Yes/No)
6	Polyphagia	Excessive hunger (Yes/No)
7	Genital Thrush	Itching or burning sensation in the genital area (Yes/No)
8	Visual Blurring	Blurred vision (Yes/No)
9	Itching	Itching on skin or body (Yes/No)
10	Irritability	Increased irritability (Yes/No)
11	Delayed Healing	Slow healing of wounds (Yes/No)
12	Partial Paresis	Partial paralysis or muscle weakness (Yes/No)
13	Muscle Stiffness	Stiffness in muscles (Yes/No)
14	Alopecia	Hair loss or balding (Yes/No)
15	Obesity	Presence of obesity (Yes/No)
16	Polydipsia	Excessive thirst (Yes/No)
17	Class	Diabetes or Non-diabetes

3.3 Hybrid SMOTE-ENN and GAN Framework for Class Imbalance

We addressed the challenge of class imbalance in the dataset by developing a hybrid data augmentation technique that combines SMOTE-ENN with Generative Adversarial Networks (GANs). This strategy was developed to effectively address the issue. The SMOTE-ENN method was initially applied to the scaled feature matrix. This was done in order to assess its

performance. In order to generate synthetic minority class samples, a technique known as the Synthetic Minority Oversampling Technique (SMOTE) is utilized. This approach involves interpolating between actual minority cases. Edited Nearest Neighbours (ENN), on the other hand, is a method that eliminates noisy or overlapping majority samples by employing a nearest neighbour cleaning procedure. This results in the elimination of the majority samples. This procedure ultimately results in a dataset that is more pristine and well-balanced than it was before. Following this, we proceeded to train a GAN using only the minority class samples that were obtained from the SMOTE-ENN output. The generator network and the discriminator network serve as integral components of the generative adversarial network (GAN). The generator network transforms random noise vectors into synthetic feature samples, while the discriminator network distinguishes between authentic data and generated counterparts. The binary cross-entropy loss technique was utilized until convergence was achieved to train both networks in an adversarial manner. Following the conclusion of the training, the generator efficiently commenced the synthesis of supplementary minority class samples. The final enhanced dataset was developed by integrating the synthetic samples produced by the GAN with the data refined through the SMOTE-ENN method. This dual augmentation strategy effectively utilizes the complementary strengths of traditional oversampling and deep learning-based data generation. This results in an improvement in the balance and diversity of the minority class, eventually improving model training outcomes. The hybrid resembling and data augmentation system utilizes SMOTE-ENN and GAN during the construction and training phases. Fig. 2 illustrates the schematic that outlines the workings of this framework.

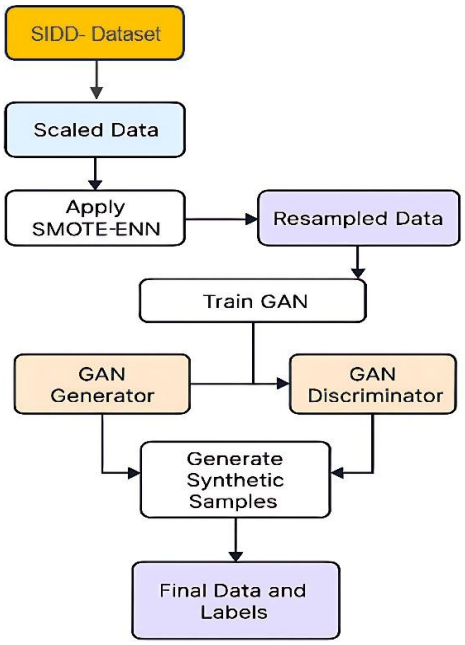


Fig 2: Proposed Hybrid SMOTE-GAN Augmentation

3.4 SHAP (SHapley Additive Explanations) Feature Selection Technique

In the last several years, the need of being able to understand machine learning has become a major issue, especially in important fields like banking, healthcare, and the legal system. Scott Lundberg and Su-In Lee came up with the SHAP (SHapley Additive exPlanations) method in 2017. This approach is acknowledged as an essential method to address this requirement. The esteemed publication "A Unified Approach to Interpreting Model Predictions" presented SHAP as a model-agnostic technique grounded in the principles of cooperative game theory. This method provided reliable and theoretically sound attributions of feature importance across different machine learning models. The SHAP algorithm employs principles from game theory, specifically Shapley values, to distribute rewards among participants in a fair and just manner. This is achieved by assigning a contribution score to each characteristic for particular predictions [24]. This framework accommodates a wide range of models, including decision trees, ensemble methods, and deep learning architectures. The effectiveness has been confirmed across multiple domains, including the identification of medical risk factors and the enhancement of model performance through feature engineering informed by interpretability [25, 26]. This research employed a Random Forest classifier to assess feature significance and model performance, utilizing the implementation offered by the scikit-learn library. The model was developed utilizing 100 estimators and a specified random seed (random_state=42) to ensure consistency in the results. Upon concluding the training phase, we utilized the SHAP framework to analyze and interpret the predictions generated by the model. A Tree Explainer was employed to compute SHAP values for the training dataset, facilitating the evaluation of the influence of each feature on the model's output. This approach provided a thorough understanding of the influence of features, enhancing the model's clarity and interpretability.

3.5 Deep Neural Networks (DNNs)

Due to their ability to understand complex and non-linear relationships through hierarchical feature representations, Deep Neural Networks (DNNs) have become essential in modern machine learning, particularly for addressing binary classification challenges. The networks consist of numerous layers, each designated to extract progressively abstract properties from the input, ultimately enhancing the model's predictive capabilities. Deep neural networks demonstrate significant efficacy in managing extensive datasets and addressing complex feature interactions, as evidenced by previous research on architecture optimization and the handling of class imbalance in binary classification [19, 20]. This study presents a DNN model that includes an input layer and three hidden layers containing 256, 128, and 64 neurons, respectively. Every layer utilizes the ReLU activation function, succeeded by Batch Normalization to enhance training stability and Dropout layers at rates of 0.4, 0.3, and 0.2 to avoid over fitting. The output layer employs a sigmoid activation function to facilitate binary classification. The model has been compiled utilizing the Adam optimizer alongside binary cross entropy as the loss function, both of which are recognized as standard selections for binary classification tasks. This design is intended to enhance learning while maintaining robustness and generalization. Fig 3 presents the proposed architecture of the Deep Neural Networks along with the corresponding flowchart. Table 1 presents the complete layer-wise architecture of the proposed Deep Neural Network (DNN) model.

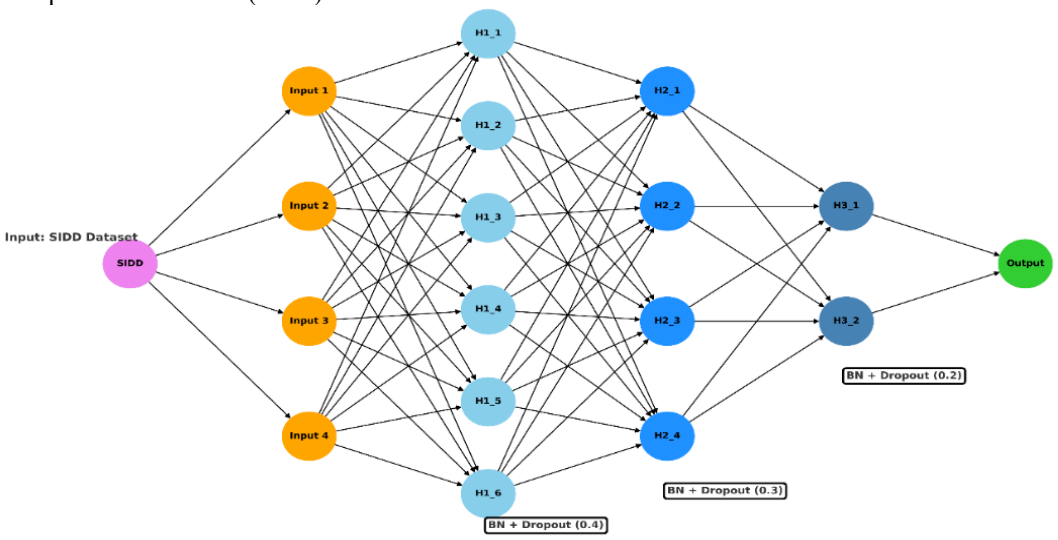


Fig 3: Proposed Deep Neural Network Architecture

Table 2: Layer-by-layer architecture of the proposed DNN model

Layer Type	Output Shape	No. of Parameters	Activation Function	Remarks
Input Layer	(input_dim,)	0	–	Receives raw features from dataset
Dense Layer	(256,)	256 × input_dim + 256	ReLU	First hidden layer
Batch Normalization	(256,)	1024	–	Stabilizes learning
Dropout Layer	(256,)	0	–	Dropout rate = 0.4
Dense Layer	(128,)	32,896	ReLU	Second hidden layer
Batch Normalization	(128,)	512	–	–
Dropout Layer	(128,)	0	–	Dropout rate = 0.3
Dense Layer	(64,)	8,256	ReLU	Third hidden layer
Batch Normalization	(64,)	256	–	–
Dropout Layer	(64,)	0	–	Dropout rate = 0.2
Output Layer (Dense)	(1,)	65	Sigmoid	Binary classification output

3.6 Radial Basis Function Neural Networks (RBFNNs)

Radial Basis Function Neural Networks (RBFNNs) are a category of feed forward neural networks that utilize radial basis functions, commonly Gaussian, as activation units in the hidden layer. RBFNNs were initially employed for function approximation tasks [21], but they are now frequently utilized for pattern recognition, classification, and time-series forecasting due to their ability to model localized receptive fields and effectively capture complex nonlinearities. A Radial Basis Function Neural Network (RBFNN) typically consists of an input layer, a singular hidden layer composed of radial basis units, and an output layer that can be either linear or logistic. A radial basis function modifies the manner in which each hidden unit evaluates the input vector against a trainable centre. RBFNNs typically demonstrate a quicker convergence rate compared to multilayer perceptrons (MLPs), particularly in scenarios where the decision boundaries are complex but spatially limited [22].A custom RBF layer was developed for this research utilizing Keras. This layer enables the application of Gaussian similarity functions to transform the input features in a nonlinear manner. Utilizing a fixed spread parameter (γ= 0.5) to regulate the Gaussian width, each RBF unit evaluates the Euclidean distance between the input vector and its centre. The proposed design includes an input layer, a 30-unit RBF layer, and an output neuron activated by a sigmoid function for binary classification. The Adam optimizer is employed to train the network comprehensively, utilizing binary cross-entropy as the loss function and accuracy as the performance metric. The architecture of Radial Basis Function Neural Networks (RBFNNs) is showed in Fig. 4. Table 3 presents the complete layer-wise architecture of the proposed Radial Basis Function Neural Networks model.

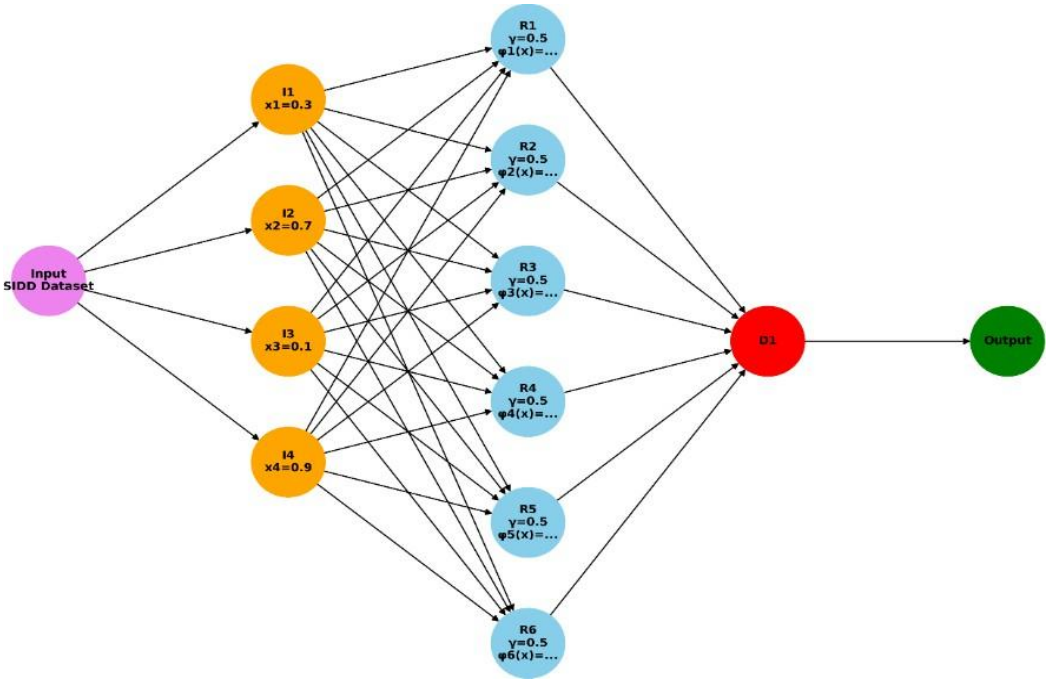


Fig 4: Proposed architecture of Radial Basis Function Neural Networks (RBFNNs)

Table 3: Layer-by-layer architecture of the proposed Radial Basis Function Neural Networks model

Layer Type	Output Shape	Number of Units	Activation	Description
Input	(input_dim,)	–	–	Accepts input feature vector
Custom RBFLayer	(None, 30)	30 (RBF Centers)	RBF Kernel	Computes Gaussian similarity with 30 trainable centers
Dense (Output)	(None, 1)	1	Sigmoid	Outputs binary classification probability

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses the outcomes of the experiments conducted using the proposed methodology, which integrates a hybrid data augmentation approach (SMOTE-ENN + GAN), SHAP-based feature selection, and neural network classification. This method aims to enhance the accuracy and clarity of diabetes risk classification through the utilization of the SIDD-1680 dataset. The original dataset exhibited significant class imbalance, which posed challenges for training and generalizing models effectively. Fig. 5 illustrates the distribution of each feature within the dataset. The distinction between the diabetes and non-diabetic samples highlighted the significance of employing a robust data balancing method to ensure that classifiers operate under equitable learning conditions. A hybrid augmentation technique was employed to address these challenges. It utilized SMOTE-ENN for oversampling and data cleaning, along with GANs to generate synthetic data. This procedure resulted in a significantly more equitable distribution of the class. Table 4 presents the count of samples prior to and following the hybrid augmentation, along with the final distribution employed for dividing the data into training and testing sets. Following the hybrid augmentation, an analysis using SHAP (SHapley Additive exPlanations) was conducted to enhance the interpretability of the features and reduce the dimensionality. The model was trained using the top 10 most significant features, determined by their SHAP values. Fig. 6 presents a SHAP summary graphic illustrating the contribution of each feature, while Fig. 7 ranks the features according to their mean SHAP significance scores derived from a Random Forest model. The model prediction was founded on essential factors such as age, delayed healing polyuria, obesity and visual blurring. The selected SHAP-based features were subsequently used to train Deep Neural Networks (DNNs) and Radial Basis Function Neural Networks (RBFNNs). We evaluated the effectiveness of their performance with and without hybrid augmentation. Fig. 8 illustrate that the model accuracy comparison, showing an improvement from 74% without augmentation to 98% with augmentation. The efficacy of the RBFNN was substantially enhanced, increasing from 73% to 94%. Table 5 presents the metrics utilized to evaluate the effectiveness of the categorization process. The measures encompass accuracy, recall, and F1-score. The DNN model achieved a macro F1-score of 98%, indicating strong performance. The performance was commendable on a class-by-class evaluation, leading to the achievement of this score. The RBFNNs performed effectively, achieving a macro F1 score of 94%. The results indicate that both models perform effectively when trained on features selected by SHAP. Fig. 9 displays a matrix of confusion of the DNN, in addition to the training and validation accuracy curves, and the ROC curve. The model exhibits strong generalization capabilities and displays minimal over fitting, supported by training and validation accuracies of 97%. The ROC curve is effective in distinguishing between different objects. Table 6 provides a detailed overview of the classification report of the models. Table 7 shows Confidence-based prediction results of DNN on the test set and in Table 8 shows Comparative assessment of a suggested model against established methodologies.

Table 4: Class-wise Data Allocation Before and After SMOTE-ENN with Train–Test Split

Class	Before SMOTE-ENN+GAN	After SMOTE-ENN+GAN	Train Set	Test Set
Diabetes	1114	1051	840	211
Non Diabetes	566	1000	800	200
Total	**1680**	**2051**	**1640**	**411**

Table 5: Accuracy, loss, and test performance of DNNs and RBFNNs with and without augmentation

Algorithms	Accuracy without SMOTE-ENN+GAN					Accuracy with SMOTE-ENN+GAN				
	Train	Loss	Val	Loss	Test	Train	Loss	Val	Loss	Test
DNNs	75%	0.4739	68%	0.5407	74%	97%	0.0542	97%	0.0346	98%
RBFNNs	73%	0.5339	77%	0.4886	73%	94%	0.1690	95%	0.1657	94%

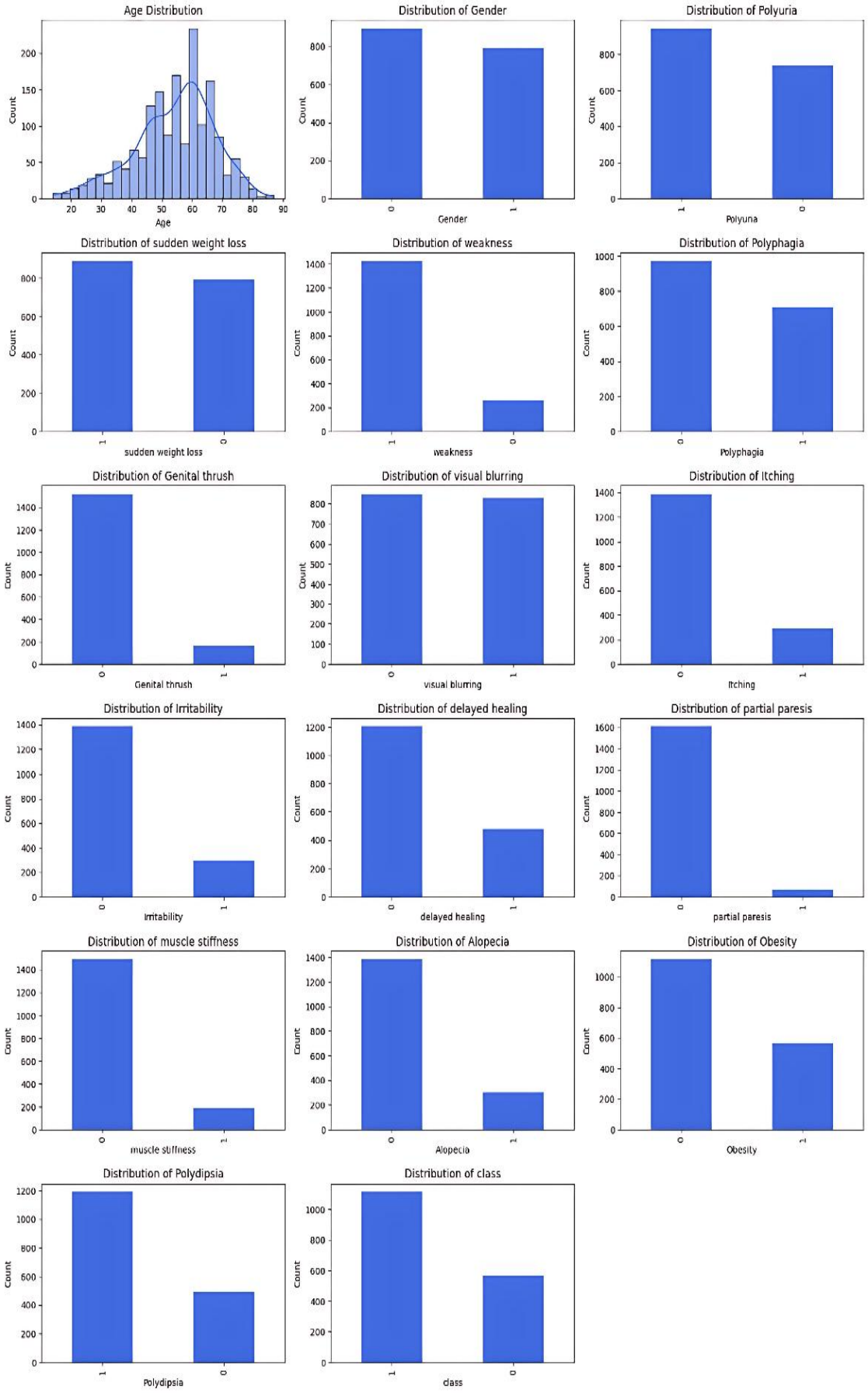


Fig 5: Distribution of each feature in the dataset before augmentation

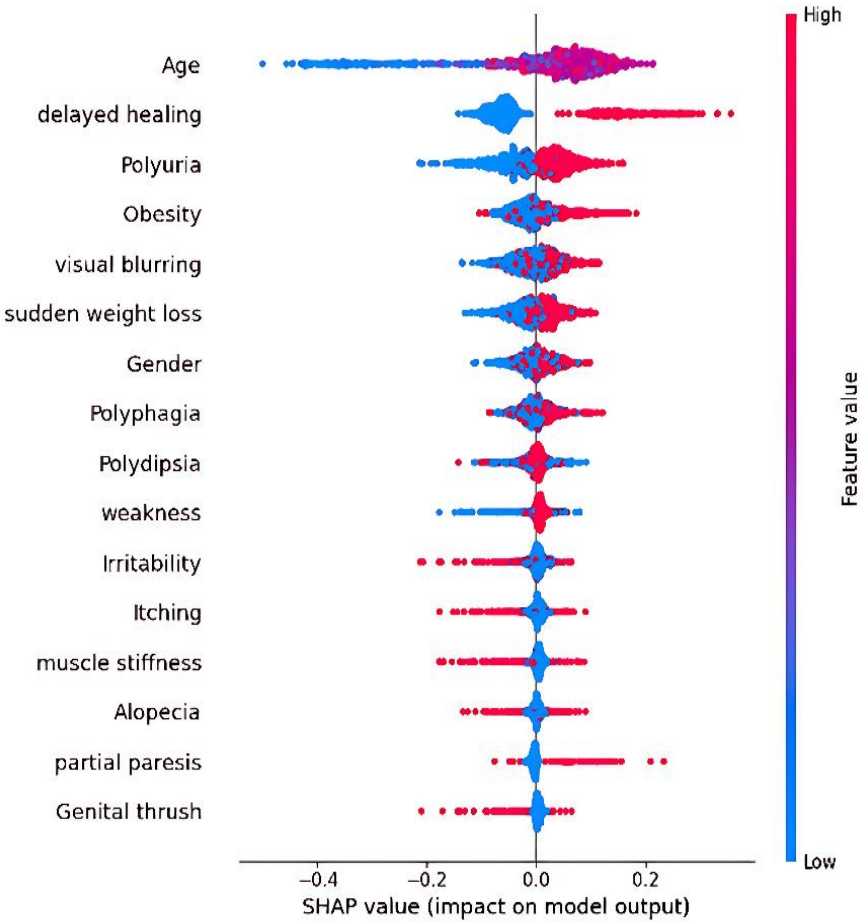


Fig 6: SHAP summary plot illustrating the contribution of each feature post-augmentation

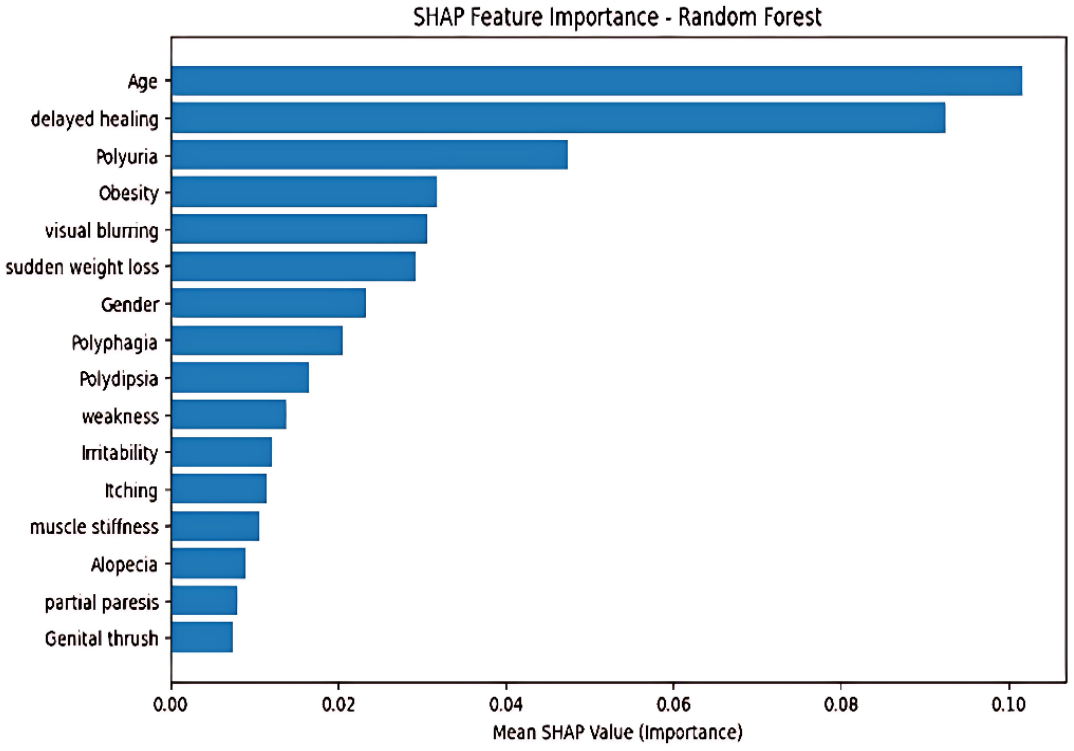


Fig 7: Mean SHAP-based feature importance ranking using Random Forest

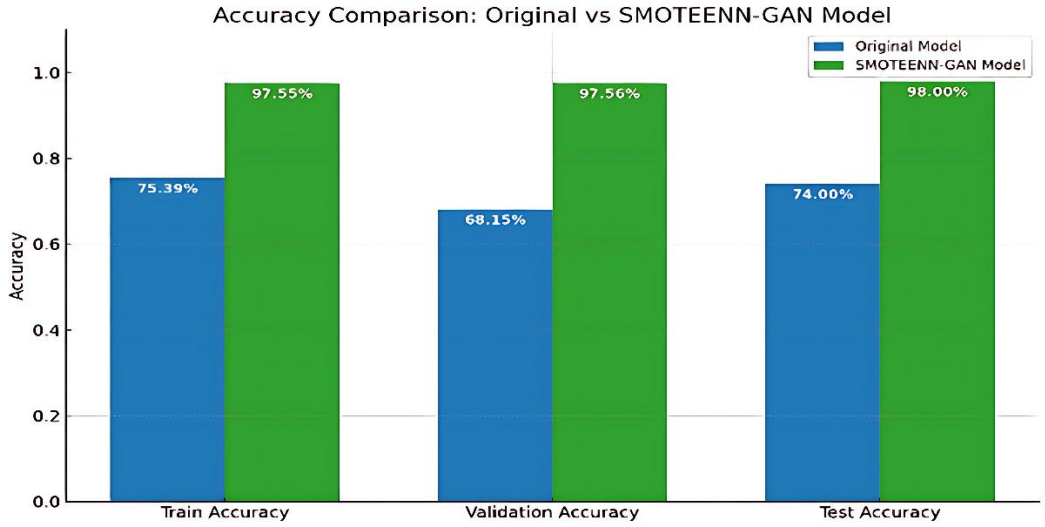
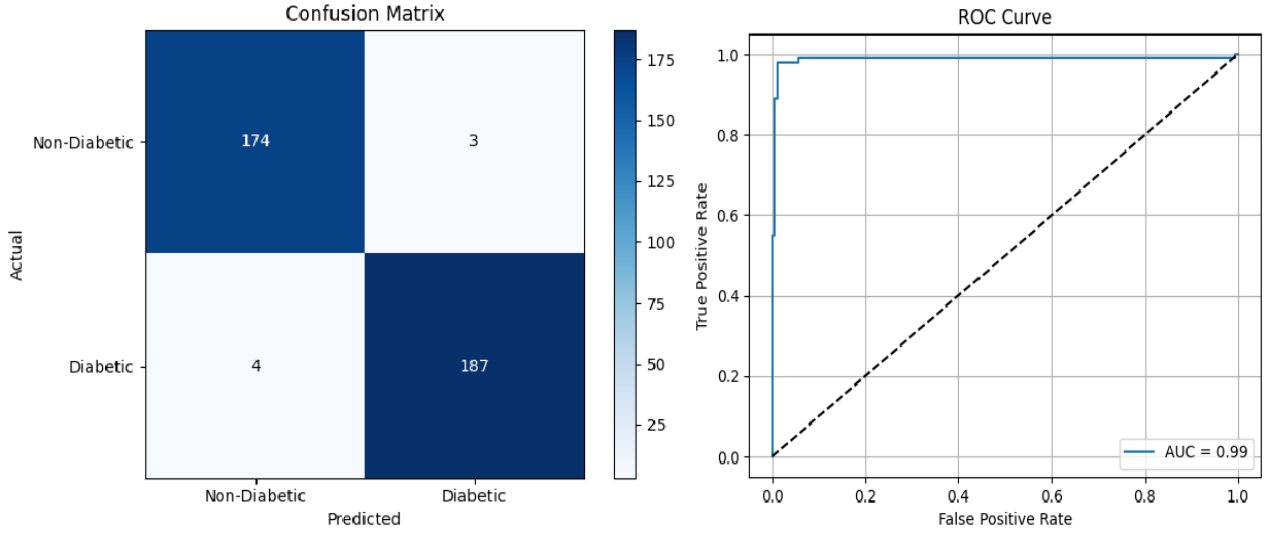
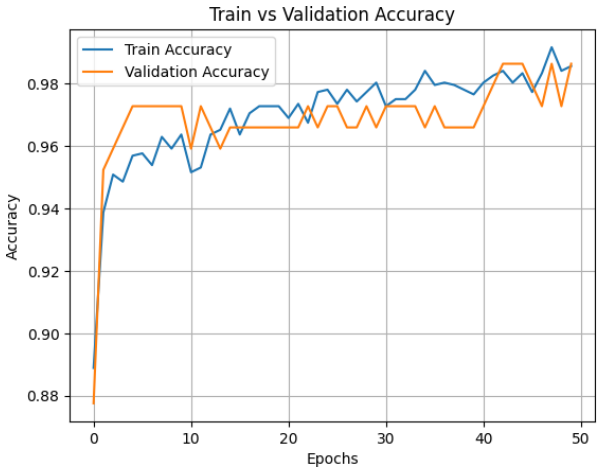


Fig 8: Model accuracy comparison before and after SMOTE-ENN+GAN augmentation



(a) **(b)**



(c)

Fig 9: DNNs model performance shown by (a) Confusion Matrix, (b) ROC Curve and (c) Train vs Validation Accuracy

Table 6: Classification report for DNN and RBFNN models using SHAP-selected features

Algorithms	Class	Precision	Recall	F1-score	Support
Deep Neural Networks	0	98 %	97%	98%	200
	1	98%	98%	98%	211
	macro avg	98%	98%	98%	411
	weighted avg	98%	98%	98%	411
Radial Basis Function Neural Networks	0	92 %	95%	94%	200
	1	95%	92%	94%	211
	macro avg	94%	94%	94%	411
	weighted avg	94%	94%	94%	411

Table 7: Confidence-based prediction results of DNN on the test set

Sl.No	Actual	Predicted	Confidence
1	1	1	0.999541
2	0	0	0.000481
3	1	1	0.999339
4	1	1	0.999868
5	0	0	0.000012
6	1	1	0.999755
7	0	0	0.000033
8	0	0	0.000122
9	1	1	0.999860
10	0	0	0.022791

Table 8: Comparative assessment of a suggested model against established methodologies

Authors	Methods	Dataset	Accuracy
Priyadarshinee, S., & Panda, M. [27]	Deep Neural Networks	PIMA Dataset	93.7%
Ashiquzzaman, A. et al [28]	Deep Neural Networks	PIMA Dataset	88.41%
Proposed Method	**Deep Neural Networks**	**SIDD-1680**	**98%**

5. CONCLUSION

This study introduces a clinically validated and interpretable deep learning framework for the early detection of Type 2 Diabetes Mellitus (T2DM), utilizing symptom-based and demographic features. A novel, region-specific dataset Southern India Diabetes Dataset (SIDD) was constructed using ethically approved clinical records from the Koppal Institute of Medical Sciences, ensuring contextual and diagnostic relevance. To overcome the inherent class imbalance, a hybrid augmentation technique combining SMOTE-ENN and Generative Adversarial Networks (GANs) was employed, resulting in a balanced and enriched training dataset. Feature selection was performed using SHAP (SHapley Additive exPlanations) values from a Random Forest classifier, enhancing model transparency and interpretability. Two deep learning architectures, Deep Neural Networks (DNN) and Radial Basis Function Neural Networks (RBFNN), were trained on the SHAP-selected features. The DNN model achieved a test accuracy of 98% with a macro F1-score of 98%, outperforming the RBFNN, which achieved 94% accuracy. All outputs underwent thorough review and validation by qualified medical experts, enhancing the clinical reliability of the framework. The findings validate the proposed approach's capability as a scalable, non-invasive screening

tool for assessing early-stage diabetes risk in healthcare settings with limited resources. The forthcoming research will focus on broadening the dataset by integrating multi-regional and multi-institutional clinical records to enhance the generalizability of the predictive models. Furthermore, the exploration of more sophisticated deep learning architectures will be undertaken to improve classification accuracy and robustness.

REFERENCES

[1] Anjana, R. M., Unnikrishnan, R., Deepa, M., Pradeepa, R., Tandon, N., Das, A. K., ... & Ghosh, S. (2023). Metabolic non-communicable disease health report of India: the ICMR-INDIAB national cross-sectional study (ICMR-INDIAB-17). The Lancet Diabetes & Endocrinology, 11(7), 474-489.

[2] Ramachandran, A., Snehalatha, C., & Ma, R. C. W. (2014). Diabetes in south-east Asia: An update. Diabetes research and clinical practice, 103(2), 231-237.

[3] Anjana, R. M., Pradeepa, R., Deepa, M., Datta, M., Sudha, V., Unnikrishnan, R., ... & ICMR–INDIAB Collaborative Study Group. (2011). Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research–INdia DIABetes (ICMR–INDIAB) study. Diabetologia, 54, 3022-3027.

[4] Anjana, R. M., Deepa, M., Pradeepa, R., Mahanta, J., Narain, K., Das, H. K., ... & Yajnik, C. S. (2017). Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR–INDIAB population-based cross-sectional study. The lancet Diabetes & endocrinology, 5(8), 585-596.

[5] Kumar, A., Gangwar, R., Ahmad Zargar, A., Kumar, R., & Sharma, A. (2024). Prevalence of diabetes in India: A review of IDF diabetes atlas 10th edition. Current diabetes reviews, 20(1), 105-114.

[6] Rooney, M. R., Fang, M., Ogurtsova, K., Ozkan, B., Echouffo-Tcheugui, J. B., Boyko, E. J., ... & Selvin, E. (2023). Global prevalence of prediabetes. Diabetes Care, 46(7), 1388-1394.

[7] Ortiz-Martínez, M., González-González, M., Martagón, A. J., Hlavinka, V., Willson, R. C., & Rito-Palomares, M. (2022). Recent developments in biomarkers for diagnosis and screening of type 2 diabetes mellitus. Current diabetes reports, 22(3), 95-115.

[8] Zhang, X., Qian, B., Li, Y., Cao, S., & Davidson, I. (2021). Context-aware and time-aware attention-based model for disease risk prediction with interpretability. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3551-3562.

[9] Shahin, O. R., Alshammari, H. H., Alzahrani, A. A., Alkhiri, H., & Taloba, A. I. (2023). A robust deep neural network framework for the detection of diabetes. Alexandria Engineering Journal, 74, 715-724.

[10] Prabhu, P., & Selvabharathi, S. (2019, July). Deep belief neural network model for prediction of diabetes mellitus. In 2019 3rd international conference on imaging, signal processing and communication (ICISPC) (pp. 138-142). IEEE.

[11] Panigrahy, S., Dash, S., & Padhy, S. (2024). Optimized Deep Belief Networks Based Categorization of Type 2 Diabetes using Tabu Search Optimization. International Journal of Advanced Computer Science & Applications, 15(3).

[12] Lang, L. Y., Gao, Z., Wang, X. G., Zhao, H., Zhang, Y. P., Sun, S. J., ... & Austria, R. S. (2021). Diabetes prediction model based on deep belief network. Journal of Computational Methods in Science and Engineering, 21(4), 817-828.

[13] Liu, Y., Zhao, Z., Wang, J., Li, A., & Zhang, J. (2019). Research on Diabetes Management Strategy Based on Deep Belief Network. In Wireless and Satellite Systems: 10th EAI International Conference, WiSATS 2019, Harbin, China, January 12–13, 2019, Proceedings, Part II 10 (pp. 177-186). Springer International Publishing.

[14] Reddy, S. S., Sethi, N., & Rajender, R. (2020, July). Evaluation of deep belief network to predict hospital readmission of diabetic patients. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 5-9). IEEE.

[15] Ma, S., Yusoh, M., & Pongsiri, N. (2023). Predicting Diabetes using Deep Belief Network. Pridiyathorn Science Journal, 2(1), 24-32.

[16] Olabanjo, O., Wusu, A., & Mazzara, M. (2023). Deep Unsupervised Machine Learning for Early Diabetes Risk Prediction using Ensemble Feature Selection and Deep Belief Neural Networks.

[17] Nadesh, R. K., & Arivuselvan, K. (2020). Type 2: diabetes mellitus prediction using deep neural networks classifier. International Journal of Cognitive Computing in Engineering, 1, 55-61.

[18] Ashiquzzaman, A., Tushar, A. K., Islam, M. R., Shon, D., Im, K., Park, J. H., ... & Kim, J. (2018). Reduction of overfitting in diabetes prediction using deep learning neural network. In IT Convergence and Security 2017: Volume 1 (pp. 35-43). Springer Singapore.

[19] Munkhdalai, L., Munkhdalai, T., & Ryu, K. H. (2020). GEV-NN: A deep neural network architecture for class imbalance problem in binary classification. Knowledge-Based Systems, 194, 105534.

[20] Alparslan, Y., Moyer, E. J., Isozaki, I. M., Schwartz, D., Dunlop, A., Dave, S., & Kim, E. (2021, July). Towards searching efficient and accurate neural network architectures in binary classification problems. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[21] Lowe, D., & Broomhead, D. (1988). Multivariable functional interpolation and adaptive networks. Complex systems, 2(3), 321-355.

[22] Moody, J., & Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. Neural computation, 1(2), 281-294.

[23] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[24] Van den Broeck, G., Lykov, A., Schleich, M., & Suciu, D. (2022). On the tractability of SHAP explanations. Journal of Artificial Intelligence Research, 74, 851-886.

[25] Sun, J., Sun, C. K., Tang, Y. X., Liu, T. C., & Lu, C. J. (2023, July). Application of SHAP for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. In Healthcare (Vol. 11, No. 14, p. 2000). MDPI.

[26] Arslan, Y., Lebichot, B., Allix, K., Veiber, L., Lefebvre, C., Boytsov, A., ... & Klein, J. (2022, August). Towards refined classifications driven by shap explanations. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 68-81). Cham: Springer International Publishing.

[27] Priyadarshinee, S., & Panda, M. (2024, March). Optimizing Diabetes Risk Prediction: Metaheuristic-Driven Tuning of Deep Neural Networks. In 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) (pp. 1-6). IEEE.

[28] Ashiquzzaman, A., Tushar, A. K., Islam, M. R., Shon, D., Im, K., Park, J. H., ... & Kim, J. (2018). Reduction of overfitting in diabetes prediction using deep learning neural network. In IT Convergence and Security 2017: Volume 1 (pp. 35-43). Springer Singapore.

[29] Early Stage Diabetes Risk Prediction [Dataset]. (2020). UCI Machine Learning Repository. DOI: 10.24432/C5VG8H.