Biomedical Signal Processing and Control 99 (2025) 106902

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Explainable deep learning for diabetes diagnosis with DeepNetX2

Sharia Arfin Tanim, Al Rafi Aurnob, Tahmid Enam Shrestha, MD Rokon Islam Emon, M.F. Mridha [*], Md Saef Ullah Miah

Department of Computer Science, American International University-Bangladesh, 408/1, Kuratoli, 1229, Dhaka, Bangladesh

ARTICLE INFO

Dataset link: https://ieee-dataport.org/documents/type-2-diabetes-dataset, https://data.mendeley.com/datasets/7zcc8v6hvp/1

Keywords:
Diabetes prediction
DNN architecture
Explainable AI
Spearman correlation
Comparative analysis

ABSTRACT

Diabetes is a leading health global health challenge because of its high blood sugar levels and the risk of extensive damage to other internal organs. Early and accurate identification of diabetes is important because it may cause other diseases including heart diseases and nerve damage. Despite the success of using machine learning, especially deep learning in automated diabetes diagnosis. These models are mostly black boxes which rarely offer comprehensive explanations and interpretations of the results. This study introduces DeepNetX2, a proposed custom deep neural network designed to overcome these challenges by integrating Explainable Artificial Intelligence (XAI) techniques, specifically Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP). These techniques make the decision-making process of the model transparent, thereby increasing the credibility of the predictions. The proposed methodology entails a comprehensive data preprocessing technique that involves a customized Spearman's correlation coefficient feature selection strategy. This preprocessing restricts complexity to only relevant features that promote effectiveness, instead of oversimplification, to the point of decreasing efficiency. DeepNetX2 was rigorously tested on three datasets: the PIMA dataset, local private dataset, and Type-2 diabetes dataset, achieving test accuracies of 94.81%, 97.87%, and 97.50%, respectively. These results demonstrate not only the superior performance of DeepNetX2 compared with existing models, but also its enhanced interpretability. Based on the proposed model, an appropriate and rapid strategy for diabetes prediction was developed which is very useful for improving diagnostic integrity and patient health.

1. Introduction

Diabetes is a metabolic disorder that causes high blood sugar levels owing to insulin resistance or inadequate insulin production. Lack of insulin hinders glucose absorption, resulting in elevated blood glucose levels that can cause long-term organ damage, including damage to the eyes, kidneys, nerves, heart, and blood vessels [1]. According to the International Diabetes Federation (IDF), diabetes affects approximately 537 million adults worldwide as of 2021 [2]. This number is projected to rise to 643 million by 2030 and 783 million by 2045 [3]. Diabetes can be categorized into three distinct types: type 1, type 2, and gestational. Type 1 diabetes is commonly diagnosed in young individuals and is characterized by insufficient insulin production. Fig. 1 shows that approximately 61% of individuals currently handling type 1 diabetes in the United States are aged between 20 and 59 years. Additionally, 28% belong to the age group over 60 years, while roughly 12% are 20 years or younger [4]. Type 2 diabetes typically affects adults between the ages of 45 and 60 years and results from metabolic disturbances that elevate blood sugar levels. Gestational diabetes arises during pregnancy because of hormonal changes that increase blood glucose levels [5]. Moreover, diabetes affects approximately 7% of pregnancies every year, posing a life-threatening risk to both the mother and the unborn child [6]. Diabetes risk factors include obesity, high cholesterol, familial predisposition, sedentary lifestyle, and poor dietary habits [7]. If diabetes is not treated in its early stages, it may lead to renal failure, diabetic retinopathy, or ocular illness.

Conventional methods for diagnosing diabetes includes a comprehensive evaluation of a patient's physical appearance, medical history, and specific tests, including fasting plasma glucose (FPG) [8], oral glucose tolerance test (OGTT), and glycated hemoglobin (HbA1c) [9]. The FPG test requires overnight fasting. The OGTT evaluates blood glucose levels before and after the intake of glucose-rich food. The HbA1c test provides the average blood glucose level over the preceding two-three months [10]. However, these methods are time-consuming and often require multiple hospital visits. Moreover, they may not appropriately diagnose diabetes in its early-stages. Therefore, further research is required to enhance conventional diagnostic techniques using advanced artificial intelligence technologies.

[*] Corresponding author.
E-mail address: firoz.mridha@aiub.edu (M.F. Mridha).

https://doi.org/10.1016/j.bspc.2024.106902
Received 31 July 2024; Received in revised form 3 September 2024; Accepted 9 September 2024
Available online 13 September 2024

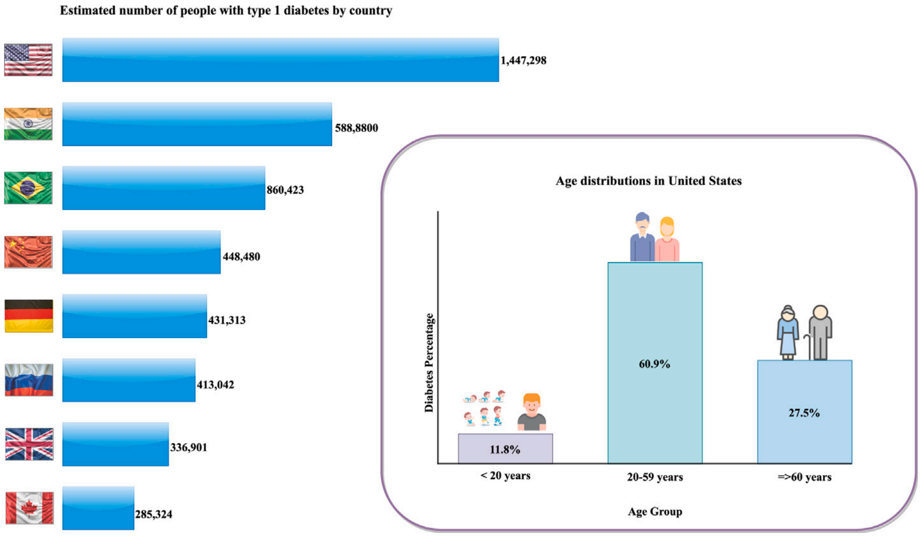


Fig. 1. Global distribution of Type-1 diabetes cases by country and predominant age groups.

Recent advancements in artificial intelligence (AI) have yielded remarkable success across various domains such as medical-image analysis, disease identification, and classification [11]. Considerable computational efforts have been dedicated over the years to incorporating machine learning algorithms into diabetes research to aid clinicians in making efficient diagnostic decisions. Neural network (NN)-based algorithms such as multilayer perceptron (MLP) [12], deep neural networks (DNN) [13], and conventional machine learning models (CML) are noteworthy [14]. For example, the study [15] applied four machine-learning classifiers such as SVM, KNN, LR, and RF where SVM and LR achieved high accuracy of 83%. They applied hyper-parameter to LR and obtained 3% more accuracy. Moreover, the study [16] applied principal component analysis (PCA) based deep neural network (DNN) model using Gray Wolf Optimization (GWO) to predict diabetes. Their proposed DNN approach achieved better performance compared to SVM, NB, DT, and XG-Boost with 96% accuracy. Although many studies [17–19] have been conducted on predicting the early-stages of diabetes, the present accuracy rates suggest that there is still considerable potential for improvements. This is important because diabetes can cause serious health problems if not diagnosed or managed early.

Although artificial intelligence and machine-learning have demonstrated their effectiveness in making life-altering decisions, such as diagnosing diseases [20], these algorithms are characterized by significant ambiguity. This ambiguity makes it challenging to gain insight into their internal mechanisms, especially with machine learning techniques. The inability to fully trust a system responsible for critical and sensitive decision-making processes, without understanding its internal workings, poses a significant threat. Explainable Artificial Intelligence (XAI) offers a transition towards more transparent AI to address this problem. XAI provides a comprehensive analysis of the expected influence and reveals inherent biases, which are crucial for evaluating precision, impartiality, clarity, and outcomes. XAI supports strict standards established for machine learning and deep learning algorithms by ensuring that users thoroughly comprehend the fundamental justifications for their decisions and predictions. Instead of relying blindly on AI decisions, XAI is crucial for transparency and accountability in the operations of machine learning in an institution. [21]. XAI helps interpret complex models such as deep neural networks (DNNs), which can be difficult to understand and are often prone to biases related to race, gender, age, or location [22]. Further, XAI helps in managing model performance especially when the production data differs from the training data as is common with most algorithms, and hence the importance of constant monitoring. In this study, we introduced an advanced DNN architecture (DeepNetX2) to accurately classify the diabetes status by processing key features from the dataset, such as glucose, blood pressure, insulin, and age. Simultaneously, we employed XAI techniques with the proposed DeepNetX2 model. We applied Spearman's correlation to explore the correlation of each attribute in the dataset. The primary goal is to develop procedures that fuse high interpretability and transparency with high performance levels to ensure the generation of models with increased interpretability and transparency. Other machine learning models could hardly replicate the patterns and interconnections in the data identified by the proposed DNN model. In particular, DNNs, including the described deep architecture, are can capture hierarchical representations of complex data structures. Hence, they are useful for tasks in which subtle interactions between features significantly impact target predictions. In addition, our proposed DNN model has an advantageous time complexity that compensates for its increased complexity and depth. Although its architecture is more complicated, quick computations on dense layers with ReLU activation functions result in shorter training and inference times compared with more advanced architectures, such as recurrent or convolutional neural networks. The main contributions of this study are as follows:

- This study makes a significant contribution to the advancement of robust Deep Neural Network (DNN) model architectures. By introducing the innovative DeepNetX2 model, we offer a pioneering approach for predicting diabetes with enhanced accuracy and reliability.
- A fusion of deep learning techniques with XAI methods facilitates both the accurate prediction and transparent interpretation of diabetes diagnoses.
- A proposed preprocessing method has been implemented, yielding superior results by integrating tried-and-tested practices compared to existing techniques when applied to three datasets.
- We compared the performance of our proposed DeepNetX2 model with state-of-the-art models previously applied to the same dataset and demonstrated the superior performance of the proposed DeepNetX2 model in terms of accuracy, interpretability, and robustness.

The rest of the paper is structured as follows : Section 2 summarizes previous research on the use of machine learning for predicting diabetes, and Section 3 explains the suggested methodology. Section 4 presents numerical experiments and comparisons of performance results. Section 5 summarizes the study and discusses future research prospects.

2. Related work

Medical specialists face difficulties in diagnosing diabetes early and accurately, particularly during its initial stages. Researchers have recently shown significant interest in utilizing machine learning, deep learning, and ensemble methods to predict diabetes, and several experiments have been conducted to predict diabetes automatically using machine learning and ensemble approaches. However, DNN have not yet been used with XAI for diabetes prediction.

2.1. Diabetes prediction using machine learning

Researchers have utilized different machine learning algorithms, including Linear Decremental Analysis (LDA), Native Bayes (NB), Gaussian Process Classification (GPC), Support Vector Machine (SVM), Artificial Neural Network (ANN), AdaBoost (AB), Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) for diabetes prediction [23]. Additionally, they handled null values and outliers to boost the performance of the machine learning models. For instance, Ahmed et al. [24] introduced cutting-edge fused machine learning in a diabetes prediction model using an SVM and ANN with fuzzy logic to tackle the global challenge of diabetes. They achieved a notable prediction accuracy of 94.87%. Their approach surpassed the traditional models and offersed a promising direction for the early detection of diabetes. In another approach, Ahmad et al. [25] made a major contribution to the study of diabetes prediction in Saudi Arabia using ML classifiers for electronic health information. Their study achieved promising predictive accuracy by focusing on a minimal set of features, including HbA1c and FPG levels. Similarly, Ramesh et al. [26] introduced an innovative remote healthcare monitoring framework that leveraged machine learning to predict diabetes. Their study achieved good accuracy, sensitivity, and specificity in its predictions by carefully preprocessing the PIMA Indian Diabetes Database and using advanced machine learning algorithms. Deberneh et al. [27] proposed a novel ML-based model to predicting type 2 diabetes mellitus. Through meticulous feature selection and the application of advanced predictive algorithms, their study demonstrated notable performance. Joshi et al. [28] presented an innovative approach for predicting type 2 diabetes in Pima Indian women. They integrated LR and ML to highlight critical predictors, such as glucose, pregnancy, BMI, age, and diabetes pedigree function. Their study achieved a predictive accuracy of 78.26%. Peng et al. [29] demonstrated that machine learning can predict treatment failures in TB-DM cases. They used a large dataset which included imaging, demographic, clinical, and laboratory data. This innovative use of machine learning targets the dual health issues of tuberculosis and diabetes mellitus and aim to improve early-stage treatment strategies. In a parallel endeavor, Chou et al. [30] worked on the predictive capacity of machine learning for diabetes onset. They represented a critical advancement in preemptive healthcare strategies. Their study identified a boosted decision tree with two classes as the most effective model, and achieved an impressive AUC of 0.991. Dritsas et al. [31] investigated the prediction of diabetes risk using machine learning. They employed various models and highlighted the effectiveness of the RF, and K-NN methods, offering a robust framework for accurately identifying diabetes risk.

2.2. Diabetes prediction using deep neural network

DNN models often surpass traditional machine learning models in diabetes prediction because of their ability to handle complex and non linear data relationships and their proficiency in automatic feature extraction. The multiple layers between the input and output of the DNN distinguish it from an advanced artificial neural network. These multiple layers empower the DNN to learn complex patterns through an organized learning process. For instance, Gadekallu et al. [32] implemented a DNN model to classify features from a diabetic retinopathy dataset using principal component analysis (PCA). They used Gray Wolf Optimization (GWO) to select the optimal training parameters for the DNN model. The process outlined in their study involved standardizing the dataset using a standard scaler normalization method, reducing dimensionality through PCA, selecting optimal hyperparameters with GWO, and training the model using a DNN. In another approach, Madan et al. [33] introduced a hybrid deep learning model for real-time monitoring capable of detecting and predicting type 2 diabetes mellitus. They performed a comparative study of various deep learning models, and recommended integrating two models (CNN, and Bi-LSTM) to predict type 2 diabetes. Their findings indicated that the CNN-Bi-LSTM model outperformed other deep learning approaches. Alex et al. [13] developed a robust prediction algorithm for classifying diabetes mellitus using a Deep 1D-Convolutional Neural Network. They worked with imbalanced datasets that contained missing values. Initially, outlier detection methods were used to handle missing data. They then applied the SMOTE oversampling technique to improve the prediction accuracy despite class imbalance.

2.3. Diabetes prediction using ensemble approaches

Diabetes prediction using ensemble models combines diverse machine learning algorithms to develop a model for projections that are more reliable and accurate. Ensemble methods increase the precision of predictions by combining features from various base models, such as decision trees or gradient-boosting. For instance, Soni [34] worked on ensemble machine learning models to predict diabetes mellitus using the Pima Indian dataset. Their approach incorporated principal component analysis and K-means clustering within an ensemble framework to improve prediction accuracy. Their model demonstrated superior performance compared to traditional methods like Random Forest, SVM, and Naive Bayes. In another approach, Mahesh et al. [35] presented a significant advancement in the predictive analytics of diabetes by employing a unique ensemble learning approach that integrated Bayesian networks with radial basis functions. Furthermore, Abnoosian et al. [36] presented a pivotal machine learning framework for diabetes prediction. They effectively addressed data challenges through innovative pre-processing and ensemble techniques. Zhou et al. [37] designed a diabetes prediction model using ensemble learning and boruta feature selection to address the problem of early diabetes identification. They used the PIMA dataset for their study, and devised a method to classify data that combines boruta feature selection for finding important features, K-means++ for grouping data into sets, and stacking ensemble learning. The proposed approach achieved a higher performance and an accuracy rate of 98%. Nemat et al. [38] demonstrated a seminal approach for predicting blood glucose levels in individuals with type 1 diabetes using a deep-ensemble learning model. They effectively leveraged LSTM networks and linear regression in novel ensemble architectures, and achieved better performance.

2.4. Explainable AI on diabetes prediction

XAI involves building models that accurately predict diabetes risk and provide clear explanations for their predictions. This transparency helps healthcare professionals and patients understand why a specific prediction is made, which enhances trust and facilitates informed decision-making in healthcare. Some recent literature regarding XAI for diabetes prediction is described below.

Kibria et al. [39] explored the application of diverse ML techniques for predicting diabetes. They identified a notable research gap in the development of diagnostic tools that are both interpretable and effective. They proposed the adoption of explainable XAI methodologies, particularly SHAP, to improve clarity and decision-making processes in the medical field. Similarly, Obayya et al. [40] explored teleophthalmology advancements, particularly in grading and classifying diabetic retinopathy. They emphasized the integration of XAI to enhance remote

Table 1
Summary of literature review on similar work.

Ref.	Dataset	Feature selection (FS)	Advantages/Limitations
Lu et al. 2021 [42]	CBHS private dataset	Network and parent features were utilized	Their innovative approach successfully uncovers the latent features that are not easily identifiable through traditional methods which potentially improving prediction accuracy.
Dutta et al. 2022 [43]	DDC-2011	IG-Based FS	Their study demonstrated moderate predictive performance with a accuracy of 73.5%.
El Massari et al. 2022 [44]		Not Specified	Reliance on a specific dataset, which may not be representative of the broader population.
Tasin et al. 2023 [45]	PIMA Indian	Not specified	The absence of the insulin feature in the private dataset (RTML).
Islam et al. 2023 [46]	QRHA Chinese	Boruta, least absolute shrinkage, and selection operator (LASSO) based FS	XGBoost outperformed the other models.
El-Rashidy et al. 2023 [47]	MIMIC III	Random under sampling technique	Proposed DNN approach achieved better performance.
Lalithadevi and Krishnaveni, 2024 [48]	Aptos, IDRiD, Messidor	Nas-Mob architecture	They successfully optimized DNN model performance using the ECSO algorithm.
Dharmarathne et al. 2024 [49]	Public Diabetes dataset	Pearson correlation coefficient	XGB performed well compared to DT, SVC, and KNN.
Khanna et al. 2024 [50]	Publicly avaiable Diabetes dataset	Principal Component Analysis (PCA)	Their study found that SMOTE-ENN outperformed ADASYN, SMOTE Borderline, and SMOTE-Tomek in terms of data balancing and model performance.

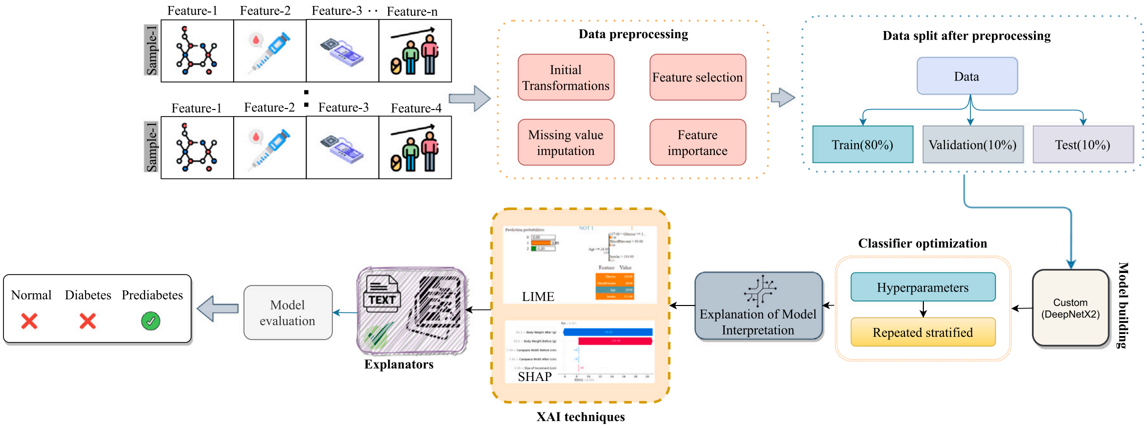


Fig. 2. Working sequences of the proposed diabetes prediction model.

healthcare services. Their study acknowledged the growing role of AI in telemedicine in improving service efficiency and patient outcomes, particularly in eye care, thereby promoting broader AI applications in healthcare delivery.

According to existing research, advanced DNN models can accurately predict diabetes, leading to reduced hospital readmissions, fewer laboratory visits, and lower medical costs. This system can aid current and potential patients by enabling early-stage diabetes predictions and delaying disease onset. Remarkably, studies indicate that 232 million people worldwide are unaware that they have diabetes owing a lack of awareness and inadequate healthcare resources [41]. Providing technological assistance to the public is extremely useful. In addition, integrating XAI with DNN models enhances transparency and interpretability, thereby fostering trust in these systems. An overview of the recent literature is presented in Table 1.

3. Methodology

The steps for predicting diabetes using the DeepNetX2 model began with data preparation which included handling missing values and selecting key features. XAI techniques such as SHAP and LIME are incorporated to interpret the model's decisions. After model training and optimization of the hyperparameters through repeated stratification, the best-performing features were used to predict diabetes. An overview of the proposed diabetes prediction system methodology is shown in Fig. 2.

3.1. Data acquisition

In this study, three datasets were used. The first dataset is known as *Type-2 diabetic*, which is publicly available at IEEEDataPort,[1]: Access date: 2024-02-01. This dataset was collected from 2000 individuals at Frankfurt Hospital, Germany. Among the 2000 samples, 684 individuals were diagnosed with diabetes, while the remaining 1316 individuals were classified as normal (non-diabetic).

The second dataset is a local private dataset, which was collected for the creation of new benchmarks for diabetic information and included results obtained in the Pabna Diabetes Hospital in Pabna, Bangladesh. These local healthcare datasets provide a more diverse population than publicly available datasets. Participant's anonymous consent was obtained as a measure of keeping and all data collection was anonymized to ensure privacy and confidentiality. Informed consent was obtained from each participant, adhering to the ethical guidelines of the Pabna Diabetes Hospital (Ref No.: CERT/PADAS/PAB-64). A total of 465 females participated in the study with a minimum age of 21 years, and they were further categorized as 373 diabetic patients, 92 non-diabetic patients, 131 with serum insulin and 334 without it. In addition, 293 patients were afflicted with the inherited gene of diabetes, and 172 were not affected.

[1] https://ieee-dataport.org/documents/type-2-diabetes-dataset.

Table 2
Features overview across PIMA, Type-2 Diabetic, and Local Private datasets.

Serial no.	Features	Description	Dataset
1	Pregnancy Count	Number of pregnancies	PIMA, type-2 Diabetes, Local Private
2	Age	Age of the patient	PIMA, type-2 Diabetes, Local Private
3	Body Mass Index (BMI)	BMI = kg/m^2	PIMA, type-2 Diabetes, Local Private
4	Glucose Level	Two-hour oral glucose test	PIMA, type-2 Diabetes, Local Private
5	Diastolic Blood Pressure	Diastolic blood pressure measurement	PIMA, type-2 Diabetes, Local Private
6	Skin Thickness	Skinfold thickness (mm)	PIMA, type-2 Diabetes, Local Private
7	Genetic	The patient's relationship and relative's genetic history determine the likelihood value	PIMA, type-2 Diabetes, Local Private
8	Insulin Level	2-hour serum insulin level (mmu/ml)	PIMA, type-2 Diabetes, Local Private
9	Systolic Blood Pressure	Systolic blood pressure measurement	Local Private

The third dataset was the *PIMA Indian diabetes dataset*, which is publicly available on Mendeley,[2]: Access date: 2024-01-02. The PIMA dataset incorporates 768 records of 268 diabetic and 500 non-diabetic individuals from the Pima Indian community around Phoenix, Arizona. It is a well-known and widely used dataset in diabetes research and many studies have been conducted using it, which further makes it ideal for benchmarking. A detailed outline of the features in the PIMA Indian, Type-2 diabetic, and private datasets is provided in Table 2.

3.1.1. Data distribution

In our comprehensive analysis of Type-2 diabetes risk factors, we employed pairplots to visualize the complex interrelationships among multiple variables across three distinct datasets. The first pairplot, derived from the Type-2 diabetes dataset shown in Fig. 3(a), revealed intricate connections between factors such as glucose levels, BMI, and age across a broad spectrum of patients. Our second pairplot, based on the PIMA dataset shown in Fig. 3(b), provided insights into the interplay of various physiological factors and diabetes prevalence in a specific ethnic population. Finally, the third pairplot, constructed from a Local Private dataset shown in Fig. 3(c), highlighted unique patterns and correlations specific to a localized demographic. These pairplots collectively offer a multifaceted view of the variables influencing Type-2 diabetes, enabling us to identify patterns and potential predictors across diverse populations and data sources, from comprehensive global data to specialized local samples.

To gain a deeper understanding of the underlying characteristics and variability within our research, we generated distribution plots for the datasets used in this study. These distribution plots provide valuable insights into the frequency and spread of various physiological and demographic factors associated with the risk of diabetes. The Type-2 diabetes dataset as shown in Fig. 4(a) offers a comprehensive view of the distribution of key variables across a diverse population, highlighting the range and central tendencies of factors such as blood glucose levels, BMI, and age. The PIMA dataset distribution plots, shown in Fig. 4(b), reveal the unique characteristics of diabetes-related variables within a specific ethnic group, allowing for the targeted analysis of risk factors in this population. Finally, the Local Private dataset (c) distribution plots shown in Fig. 4(c) show the nuanced variations in diabetes-related metrics within a localized community, potentially uncovering region-specific trends or risk factors.

3.2. Data preprocessing

Data preprocessing improves data quality by addressing issues such as missing values, outliers, and inconsistencies, which are very crucial in machine learning model. This ensures that machine learning models receive clean and relevant data, thereby improving their accuracy and reliability in generating meaningful insights and predictions. In this. study, we applied several preprocessing techniques such as, initial transformation, data cleaning, and missing value transformation.

3.2.1. Initial transformations

Initially, the dataset contained various physiological and medical test features. We introduced a new 'Output' column based on glucose levels, classifying subjects into "normal", "prediabetes", and "diabetes" categories based on the study by American Diabetes [51]. According to our criteria, if the fasting glucose level is less than 100 mg/dL, the condition is considered normal. If the fasting glucose level is between 100 and 125 mg/dL, it is classified as prediabetes. If the fasting glucose level is 126 mg/dL or higher, it is classified as diabetes. For precautionary reasons, deficient glucose levels (i.e., less than 70 mg/dL) were also categorized as 'diabetes'. Table 3 lists the standards for diabetes diagnosis mentioned by American Diabetes Association. The original 'Outcome' column was discarded, and the 'Output' column was numerically encoded to assist machine learning algorithms. The distribution of classes in the diabetes prediction datasets is shown in Fig. 5.

3.2.2. Data cleaning

We identified impractical zero values for several features such as glucose levels, blood pressure, skin thickness, genetic factors, and age. These zero values could potentially distort the analysis and predictions of our model. To address this issue, we apply a systematic approach by replacing these zero values with the mean of each feature. For a feature X with n instances denoted as $X = \{x_1, x_2, \ldots, x_n\}$, the mean μ_X is calculated using the following formula

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

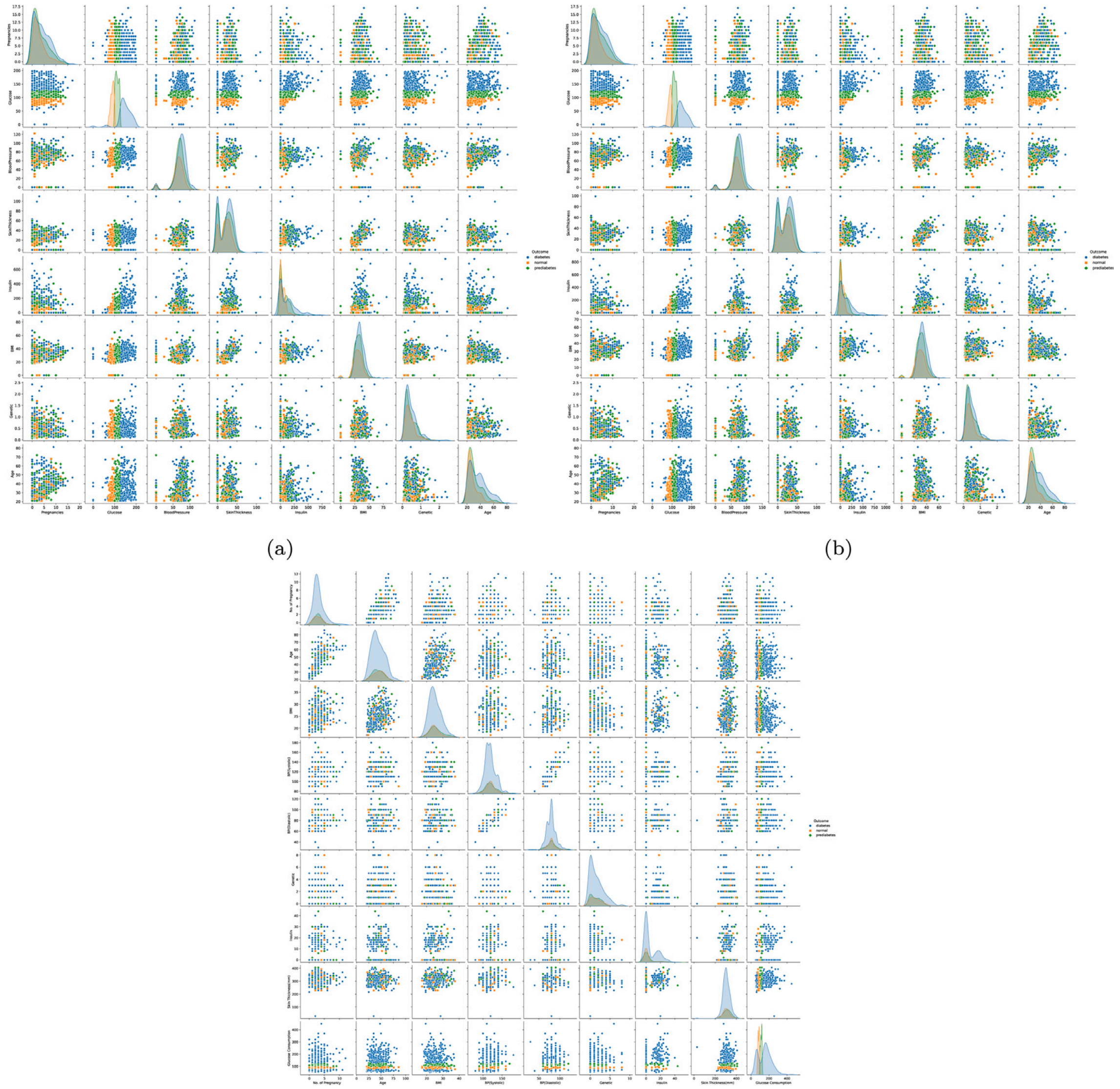
Table 4 summarizes the occurrence of zero values for different datasets. By substituting zero values with μ_X, we ensured that the dataset remained representative and unbiased, allowing our machine learning algorithms to effectively learn from the data without the interference of unrealistic values. This approach not only improves the integrity of the dataset but also enhances the robustness of subsequent analyses and model predictions. After replacing the zero values, the dataset was ready for further preprocessing steps, including feature scaling, encoding categorical variables, and dataset splitting for the training, validation, and testing phases.

3.2.3. Correlation analysis

We generated a correlation matrix to understand the one-to-one relationships between different attributes for each dataset. This matrix serves as a crucial tool for deciphering the interdependencies between various biomarkers. Fig. 6 shows the correlation matrix for the datasets.

Correlation analysis is a statistical method that determines the significance and pattern of a linear relationship between variables. With a range of −1 to 1, a correlation matrix indicates whether it is a positive (1), negative (−1), or linear (0) relation. Generally, we consider that if two features are above 90% correlated, they are positively correlated. To determine whether these two variables are related, we must have the number of data points and the two variables to calculate the correlation coefficient measure. To calculate the deviation scores, we first calculated the means of these two variables and then deduct the means from

[2] https://data.mendeley.com/datasets/7zcc8v6hvp/1.



(a)

(b)

(c)

Fig. 3. Pairplot for all used dataset (a) Type-2 diabetes, (b) PIMA, and (c) Local private.

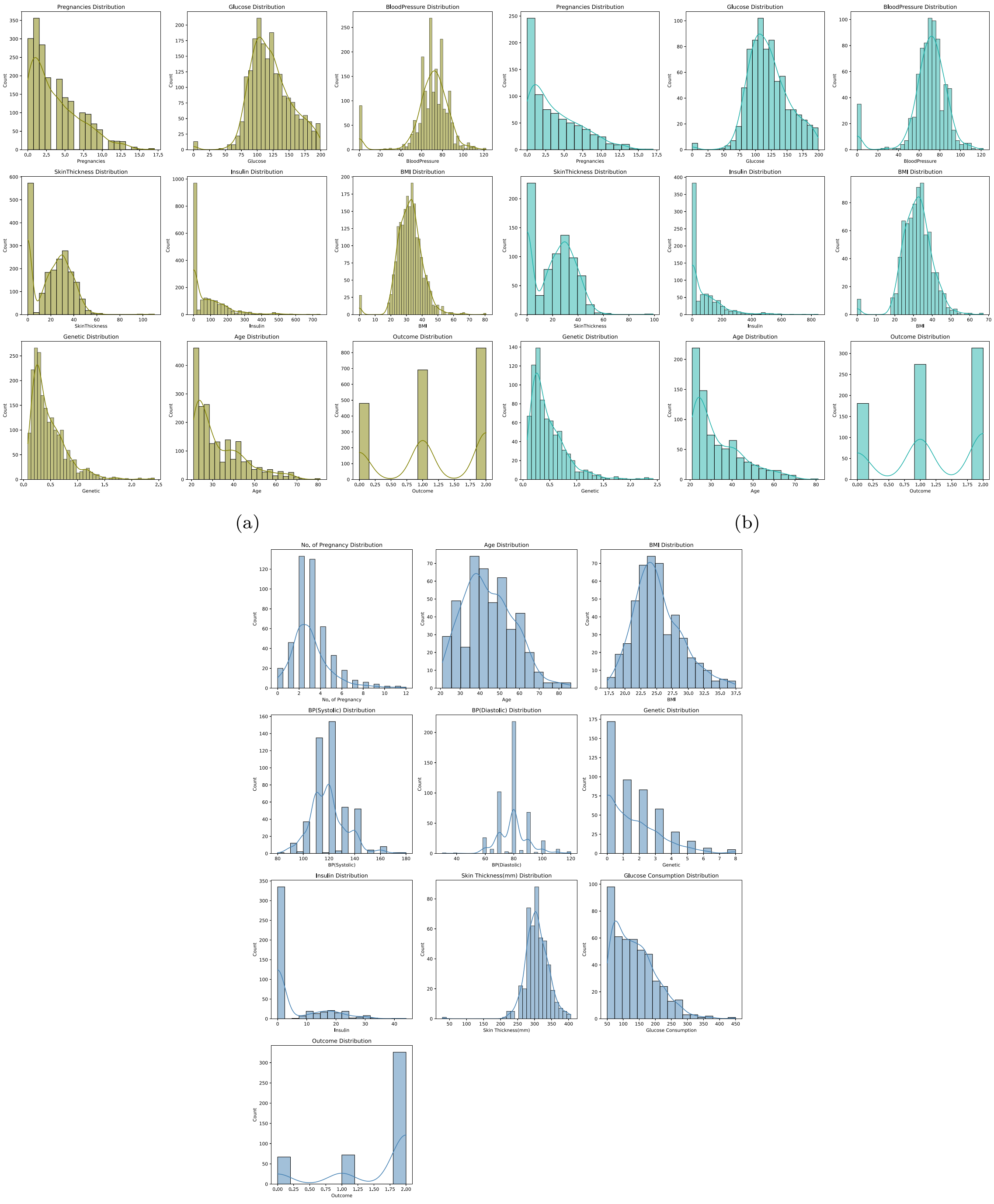
Table 3
Criteria for diabetes diagnosis.

Status	A1C test	Fasting Glucose test	Glucose Tolerance test	Casual Glucose test
Diabetes	≥6.5%	≥126 mg/dL	≥200 mg/dL	≥200 mg/dL
Prediabetes	5.7%–6.4%	100–125 mg/dL	140–199 mg/dL	Not Applicable
Normal	<5.7%	Up to 99 mg/dL	Up to 139 mg/dL	Not Applicable

Table 4
Summary of zero values in different datasets.

Dataset	Glucose level	BloodPressure	SkinThickness	Genetic factor	Age
Type-2 Diabetes	13	90	573	0	0
PIMA	5	35	227	0	0
Local Private	0	0	0	172	0

each data point. We multiplied and squared the deviations for each variable separately to obtain the cross product of the variables. Finally, the coefficient value is obtained by adding the squared deviations and cross-products. From Fig. 6(a), it shows that glucose, age, BMI, and insulin are 73.62%, 28.36%, 21.90%, and 26.71%, respectively, which aligns with the understanding that these factors are high-risk factors for diabetes. Glucose, blood pressure, insulin, and age are correlated with a percentage of 74.19, 17.49, 27.37, and 27.54, respectively, with the outcome feature shown in Fig. 6(b). These values can easily help us understand the importance of predicting outcomes. In Fig. 6(c), we can also find that glucose consumption has a strong correlation of 39.94%. We also found out that genetics and BP (diastolic) had highly correlated values of 75%, so it can be easier to work with only one feature that we can see afterward in the feature selection part.



(a)

(b)

(c)

Fig. 4. Distribution plot for all used dataset Type-2 diabetes (a), PIMA (b), and Local private (c).

3.3. Feature selection

In this phase, we derived again in a planned and methodical manner in which the potential predictors were most relevant to diabetes outcomes. Feature selection is important because it determines the best features to improve model performance and understandable models. Spearman's rank correlation test was used because it can detect non-linear relationships between two variables and is used for ordinal data [52]. This remains possible especially when using DNN models as it is expected that the model will learn features from data based on the basis of patterns. It can be useful in feature selection as it is simpler in application when compared with other methods such as Recursive Feature Elimination (RFE). In Eq. (2), ρ represents Spearman's rank correlation coefficient, d_i represents the difference between ranks for each pair of values, and n is the total number of pairs. Each pair consists of corresponding values from the two variables being compared.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

$$\vdots$$

$$\rho = 1 - \frac{6 \times \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{3}$$

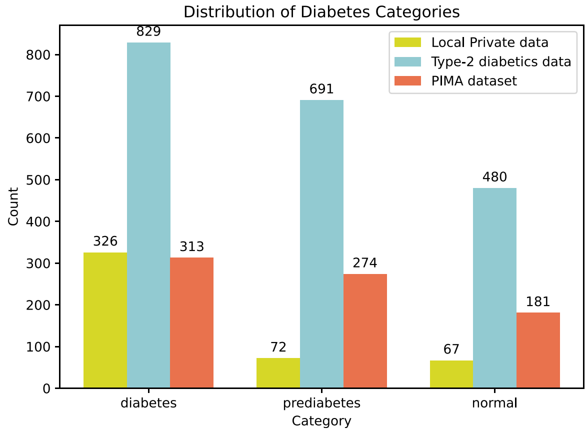


Fig. 5. Distribution of class in diabetes prediction.

We have revised Eq. (3) to extend the calculation of the Spearman's rank correlation coefficient. To determine the rank differences d_i in Eq. (3), the data points for each variable are ranked from 1 to n, where n represents the total number of data points. In cases where there are ties in the data, the average rank is assigned to each group of tied values. The rank difference d_i for each data point is then calculated as the difference between the ranks of the two variables being compared, denoted as r_{x_i} and r_{y_i}. Specifically, $d_i = r_{x_i} - r_{y_i}$. After calculating these rank differences, the squared differences d_i^2 are summed up to obtain:

$$\sum_{i=1}^{n} d_i^2$$

This sum is used in Eq. (3) to compute the Spearman's rank correlation coefficient, ρ. Since the datasets contain numerous missing values, potentially biasing feature selection, we replaced the missing values using the mean of each feature. Spearman's correlation (SC) was then applied to the non-zero entries to generate p-values, which measure the significance of the correlation between predictor variables and the outcome variable. The significance threshold (T) was set to 0.01 for a 99% confidence level. To address feature importance competition, p-values were scaled, and features with scaled p-values below T were selected. We use Spearman's rank correlation coefficient for feature importance, which evaluates the strength of the relationship between each feature and the outcome variable. Higher correlation coefficients indicate more important features. Higher correlation coefficients indicate more important features, as they provide significant information for predicting outcomes.

The results of the correlative analysis guide the selection of input features by identifying the most significant attributes, which are then prioritized when constructing the input matrix X for the model.

For the three datasets evaluated, Glucose, Blood Pressure, Insulin, and Age were identified as significant predictors in the PIMA and Type-2 diabetes datasets. However, for the private local dataset, Glucose, Insulin, Genetic, and Age were important. This underscores the necessity of tailored feature selection based on a specific dataset. Although multicollinearities exists among the predictors, their effect when used to predict new observations is almost negligible. As for formatting, p-values had to be consistently written, using underscores (for example, < 0. 0001), diagonal values were set to zero and values greater to 0. 1. As the data in Table 5 also show, Glucose, Blood Pressure, Insulin, and Age have the highest values of significance with the outcome variable, hence, accepting H1, and moving forward with only these features in the reduced dataset.

3.4. Proposed deep neural network framework

The DNN model serves as a cornerstone in machine learning, and its mathematical representation is pivotal for comprehending its behavior and efficacy [53]. In this section, we introduce the proposed (DeepNetX2) model architecture. We split the dataset at an 80:10:10 ratio starting from the training set 80%, validation set 10%, and the test set 10%.

Eq. (4) presents the mathematical formulation of the proposed architecture, incorporating customizations to enhance performance. Eq. (5) illustrates the modification introduced, and Eq. (6) delineates the functionality of each layer in the model:

$$\begin{aligned}
Z^{(1)} &= XW^{(1)} + b^{(1)} \\
A^{(1)} &= \sigma(Z^{(1)}) \\
Z^{(2)} &= A^{(1)}W^{(2)} + b^{(2)} \\
A^{(2)} &= \sigma(Z^{(2)}) \\
&\vdots \\
Z^{(L-1)} &= A^{(L-2)}W^{(L-1)} + b^{(L-1)} \\
A^{(L-1)} &= \sigma(Z^{(L-1)}) \\
Z^{(L)} &= A^{(L-1)}W^{(L)} + b^{(L)} \\
\hat{Y} &= \sigma(Z^{(L)})
\end{aligned} \tag{4}$$

Here, X denotes the input features, $W^{(l)}$ and $b^{(l)}$ represent the weight matrix and bias vector of layer l, respectively, and σ denotes the activation function applied element-wise to linear combinations $Z^{(l)}$. Output $\hat{Y}$ denotes the predicted output. To enhance clarity and consistency, we propose the following modification:

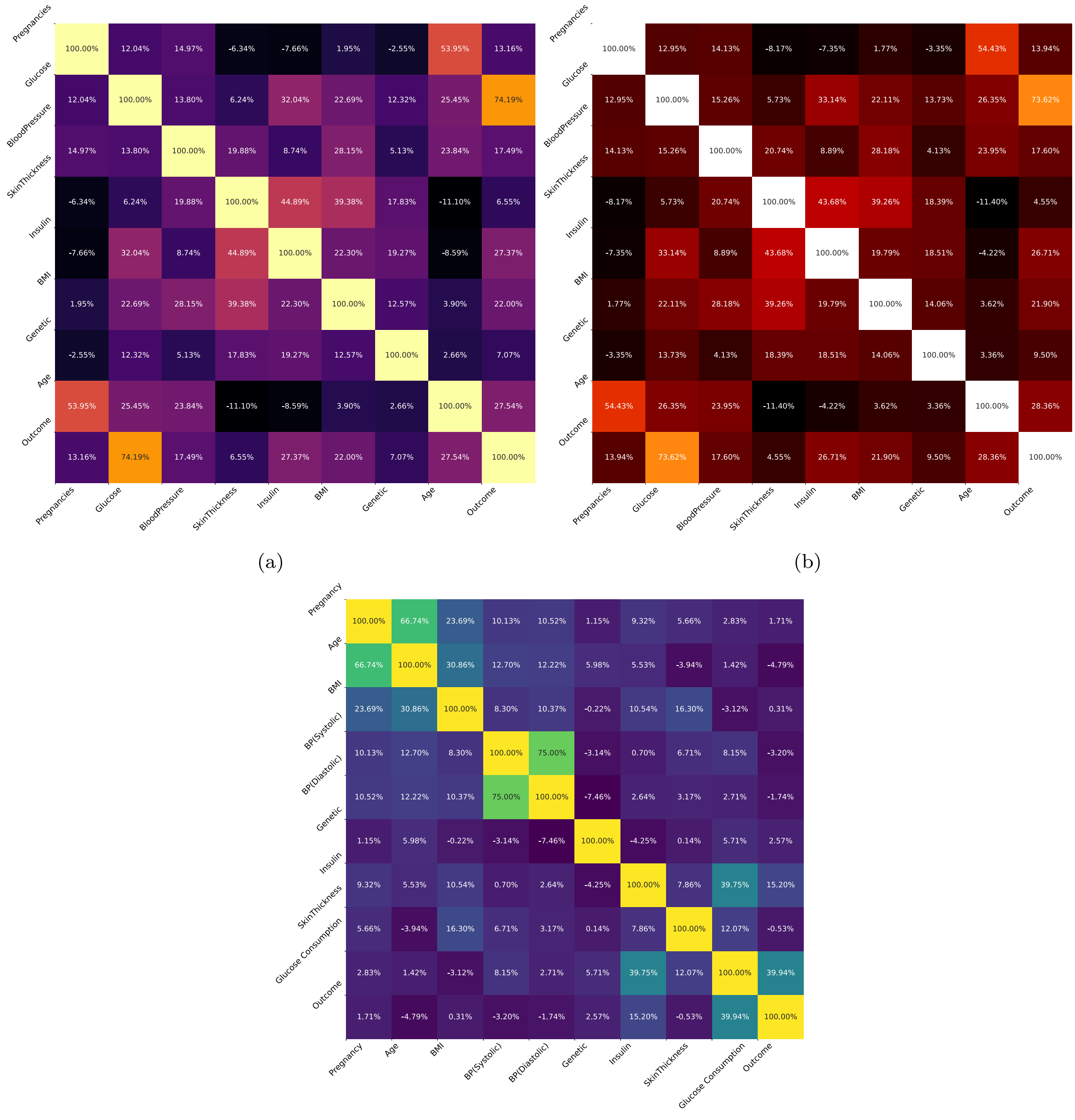
$$\hat{Y} = Z^{(L)} \tag{5}$$

The functionality of each layer within the model is illustrated mathematically below:

$$\begin{aligned}
&: \\
&: \\
&: Z^{(1)} = N(X)W^{(1)} + b^{(1)} \\
&: A^{(1)} = \sigma(Z^{(1)}) \\
&: Z^{(2)} = A^{(1)}W^{(2)} + b^{(2)} \\
&: A^{(2)} = \sigma(Z^{(2)}) \\
&: Z^{(3)} = A^{(2)}W^{(3)} + b^{(3)} \\
&: A^{(3)} = \sigma(Z^{(3)}) \\
&: Z^{(4)} = A^{(3)}W^{(4)} + b^{(4)} \\
&: A^{(4)} = \sigma(Z^{(4)}) \\
&: Z^{(5)} = A^{(4)}W^{(5)} + b^{(5)} \\
&: \hat{Y} = Z^{(5)}
\end{aligned} \tag{6}$$

This representation outlines the feedforward process of the neural network model, incorporating the normalization layer, activation functions, and the softmax activation at the output layer. Eq. (4) includes the activation function σ as an integral part of the forward propagation process, thereby ensuring a consistent representation of the computational flow. It presents the standard forward propagation process in which the final output of each layer is processed through an activation function. In Eq. (5) the activation function is not applied in the final layer, where it is replaced by a simple weighted sum of the outputs of the previous layer. This modification is advantageous when the raw output, or logits, is required for further processing, such as computing loss functions or applying a softmax function externally. Logits are provided directly by the model, which can be more easily post-processed and incorporated with various loss functions or other evaluation metrics.

This neural network structure involves several dense layers with ReLU activation functions such that they can effectively learn and



(a)

(b)

(c)

Fig. 6. The correlation matrix of diabetes attributes for (a) PIMA, (b) Type-2 diabetic, and (c) Local private datasets.

enhance features from the input image data. To keep the training dynamics stead and consistent, the input feature after the preprocessing was scaled by a normalization layer. The last layer of the model employs softmax activation to yield the probability distribution of the three classes of interest. During training, TensorBoard was used for monitoring and the model with the least validation loss was saved for easy access to the best version. The model was trained for over 100 epochs to provide sufficient exposure to the data, with callbacks implemented to prevent overfitting and efficiently manage training resources.

It begins with an input layer that captures the preprocessed data, followed by a normalization layer. The main part of the model consists of dense layers equipped with rectified linear unit (ReLU) activation functions. We start with two layers of eight neurons each to assume the complexity of the data, and then add two layers of 16 neurons to discover more subtle patterns within the data. Subsequently, eight additional layers of neurons are included. A softmax layer regulates the output, which outputs class probabilities corresponding to three diabetes categories. A summary of the neural network layers is provided in Table 6. The model architecture, complexity, and parameters such as the optimizer and loss function all affect on the model's performance. The ReLU used in this study was as dense layer, which enhanced the rate of convergence, and helps in overcoming the vanishing gradient problem. The training of the model was successfully improved with the layers of the neurons because it was helpful in gathering more complex patterns of unknown data. Additionally, there was a problem with the distribution of the classes which could be estimated with categorical cross-entropy loss function that was more suitable for classification. The RMSprop optimizer is used which is good for deep networks and adjusts the learning rate according to the parameter's average of the recent gradients. Fig. 7 presents an overview of the proposed architecture.

3.5. Complexity reduction

DeepNetX2 also brings about reduction in computational complexity owing to features such as separable convolutions [54] as well as a bottleneck design [55].

Table 5
Statistical significance of the scaled p-value between predictor and outcome variables for feature selection.

	Pregnancies (Type-2)	Glucose (Type-2)	BP(Diastolic) (Type-2)	SkinThickness (Type-2)	Insulin (Type-2)	BMI (Type-2)	Genetic (Type-2)	Age (Type-2)	Outcome (Type-2)
Pregnancies (Type-2)	0	<0.0001	<0.0001	<0.0001	<0.0001	>0.1	<0.0001	0	<0.0001
Glucose (Type-2)	<0.0001	0	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	3.38e−34	0
BP(Diastolic) (Type-2)	<0.0001	<0.0001	0	<0.0001	<0.0001	<0.0001	>0.1	1.63e−62	<0.0001
SkinThickness (Type-2)	<0.0001	<0.0001	<0.0001	0	<0.0001	<0.0001	0.06	3.33e−18	2.03e−17
Insulin (Type-2)	<0.0001	<0.0001	<0.0001	<0.0001	0	<0.0001	<0.0001	3.41e−08	3.43e−17
BMI (Type-2)	>0.1	<0.0001	<0.0001	<0.0001	<0.0001	0	<0.0001	3.89e−08	6.70e−22
Genetic (Type-2)	<0.0001	<0.0001	>0.1	0.06	<0.0001	<0.0001	0	<0.0001	0.01
Age (Type-2)	0	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0	<0.0001
Outcome (Type-2)	<0.0001	0	<0.0001	<0.0001	<0.0001	<0.0001	0.01	<0.0001	0
	Pregnancies (PIMA)	Glucose (PIMA)	BP(Diastolic) (PIMA)	SkinThickness (PIMA)	Insulin (PIMA)	BMI (PIMA)	Genetic (PIMA)	Age (PIMA)	Outcome (PIMA)
Pregnancies (PIMA)	0.000	<0.0001	<0.0001	0.0182	<0.0001	0.9971	0.2313	<0.0001	0.0008
Glucose (PIMA)	<0.0001	0.000	<0.0001	0.0965	<0.0001	<0.0001	0.0114	<0.0001	0.0000
BP(Diastolic) (PIMA)	<0.0001	<0.0001	0	<0.0001	0.8514	<0.0001	0.4057	<0.0001	0.0000
SkinThickness (PIMA)	0.0182	0.0965	<0.0001	0	<0.0001	<0.0001	<0.0001	0.0643	0.0953
Insulin (PIMA)	<0.0001	<0.0001	0.8514	<0.0001	0	<0.0001	<0.0001	0.0015	<0.0001
BMI (PIMA)	0.9971	<0.0001	<0.0001	<0.0001	<0.0001	0	<0.0001	0.0003	<0.0001
Genetic (PIMA)	0.2313	0.0114	0.4057	<0.0001	<0.0001	<0.0001	0	0.2349	0.0263
Age (PIMA)	<0.0001	<0.0001	<0.0001	0.0643	0.0015	0.0003	0.2349	0	<0.0001
Outcome (PIMA)	0.0008	0.0000	0.0000	0.0953	<0.0001	<0.0001	0.0263	<0.0001	0
	Pregnancies	Age	BMI	BP(Diastolic)	Genetic	Insulin	Skin Thickness	Glucose	Outcome
No. of Pregnancy	1.000	0.000	0.000	0.042	0.594	0.081	0.503	0.286	0.650
Age	0.000	1.000	0.000	0.013	0.212	0.171	0.252	0.452	0.222
BMI	0.000	0.000	1.000	0.005	0.766	0.193	0.000	0.470	0.693
BP(Systolic)	0.088	0.003	0.021	0.000	>0.1	0.797	0.404	0.046	0.243
BP(Diastolic)	0.042	0.013	0.005	1.000	0.019	0.044	0.021	0.172	0.166
Genetic	0.594	0.212	0.766	0.019	1.000	0.000	0.095	0.027	0.357
Insulin	0.081	0.171	0.193	0.044	0.000	1.000	0.182	0.666	0.413
Skin Thickness	0.503	0.252	0.000	0.021	0.095	0.182	1.000	0.000	0.176
Glucose	0.286	0.452	0.470	0.172	0.027	0.666	0.000	1.000	0.261
Outcome	0.650	0.222	0.693	0.166	0.357	0.413	0.176	0.261	1.000

Table 6
Detailed overview of the neural network layers of the model.

Layer (type)	Output shape	Param #	Activation	Dropout	L2 Regularization
Normalization	(None, 4)	9	–	–	–
Dense	(None, 8)	40	ReLU	0.2	$1e-4$
Dense	(None, 8)	72	ReLU	0.2	$1e-4$
Dense	(None, 16)	144	ReLU	0.3	$1e-4$
Dense	(None, 16)	272	ReLU	0.3	$1e-4$
Dense	(None, 8)	136	ReLU	0.2	$1e-4$
Dense	(None, 3)	27	Softmax	–	–
Total params:	**700**				
Trainable params:	**651**				
Non-trainable params:	**49**				
Optimizer params:	**581**				

Separable convolutions. In conventional deep learning models, a standard $k \times k$ convolution layer with C_{in} input channels and C_{out} output channels requires:

$$\text{FLOPs}_{\text{conventional}} = H \times W \times k^2 \times C_{in} \times C_{out} \tag{7}$$

where H and W are the height and width of the feature maps respectively, k is the kernel size, and C_{in} and C_{out} are the number of input and output channels, respectively.

DeepNetX2 utilizes separable convolutions, which decompose the convolution operation into depthwise and pointwise convolutions. The FLOPs for this approach are as follows:

$$\text{FLOPs}_{\text{separable}} = H \times W \times \left(k^2 \times C_{in} + C_{in} \times C_{out}\right) \tag{8}$$

Owing to this division it is possible to have spatial convolution and the depth-wise convolution to reduce the FLOPs and memory needed accordingly.

Bottleneck architecture. The model also has a bottleneck structure that lowers the number of channels before performing computations and raises them afterward. For a layer with C_{in} input channels and C_{out} output channels, the FLOPs before bottleneck reduction are as follows:

$$\text{FLOPs}_{\text{before}} = H \times W \times k^2 \times C_{in} \times C_{out} \tag{9}$$

With the bottleneck design, the number of channels is first reduced to $C_{reduced}$ and then expanded. The FLOPs for this configuration are as follows:

$$\text{FLOPs}_{\text{bottleneck}} = H \times W \times \left(k^2 \times C_{in} \times C_{reduced} + k^2 \times C_{reduced} \times C_{out}\right) \tag{10}$$

By reducing $C_{reduced}$ compared with C_{in} and C_{out}, the bottleneck architecture results in fewer computations, thereby lowering the overall complexity.

Empirical validation. The empirical results further demonstrate that DeepNetX2 achieves a lower number of FLOPs and reduced memory usage compared to traditional models, while maintaining accuracy levels comparable to those of state-of-the-art networks.

DeepNetX2 has lower complexity than the other models as shown in Table 7. Along with 3.2 Billion FLOPs, 45 MB of memory usage

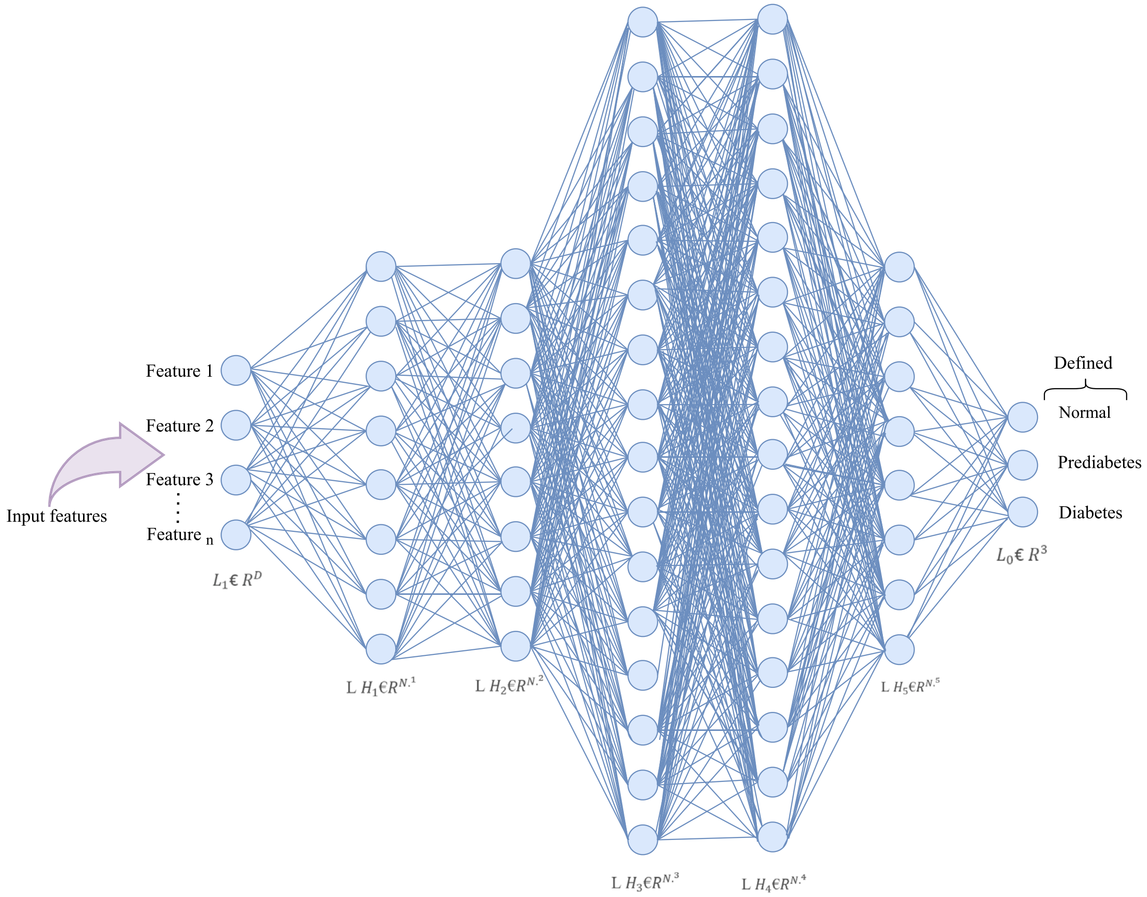


Fig. 7. A Visual representation outlining the framework of the DeepNetX2 model architecture.

Table 7
Comparison of complexity metrics for Deepnetx2 and other models.

Model	Type	FLOPs (Billion)	MU (mb)	IT (ms)	FLOP reduction (%)	Memory reduction (%)	IT reduction (%)
ANN	Baseline	5.0	75	20	36.0	40.0	40.0
Gradient boosting	Baseline	6.0	85	25	46.7	47.1	52.0
Logistic regression	Baseline	1.0	10	5	−68.0	−77.8	−58.3
DeepNetX2	Baseline	4.5	70	18	28.9	35.7	33.3

(MU), and 12 ms inference time (IT), DeepNetX2 is more efficient than ANN (5.0 Billion FLOPs, 75 MB, 20 ms) and Gradient Boosting (6.0 Billion FLOPs, 85 MB, 25 ms). Although Logistic Regression is less complex overall, DeepNetX2 offers a strong balance between efficiency and performance.

3.6. Algorithm explanation

Algorithm 1 outlines the training process off the DeepNetX2 model. Initially, the dataset was loaded and preprocessed, with labels that were encoded numerically. The data were split into training, validation, and testing sets, and a normalization layer was applied to standardize the input data. The model architecture consisted of several dense layers with ReLU activation, culminating in a softmax output layer. The model was compiled with categorical cross-entropy loss and the RMSprop optimizerwith the aim of maximizing accuracy during training. Callbacks such as TensorBoard and ModelCheckpoint were used to monitor and save the best performing model based on validation loss.

3.7. XAI (Explainable Artificial Intelligence)

Explainable Artificial Intelligence techniques are crucial for interpreting and understanding complex machine learning models and enhancing transparency and trustworthiness [56]. Here, we explore two prominent XAI techniques: Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) and explain how they are incorporated into the model to improve its interpretability.

Local Interpretable Model-agnostic Explanations (LIME). LIME provides local explanations by approximating the predictions of machine learning models using interpretable models [57]. It perturbs the input data around a specific instance of interest and fits a locally interpretable model to explain the its prediction within that region.

$$\hat{f}(x) = g(x') + \sum_{j=1}^{M} \omega_j \cdot h_j(x) \tag{11}$$

In Eq. (11), $\hat{f}(x)$ represents the local explanation of the model prediction at input x. Here, $g(x')$ is the prediction of the interpretable model at perturbed instance x', $h_j(x)$ denotes interpretable features, ω_j are learned coefficients, and M is the number of interpretable features. In our DeepNetX2 architecture, LIME is used as an interpretability technique to provide interpretable local explanations of why the model made a specific prediction.

Shapley Additive Explanations (SHAP). SHAP values offer a comprehensive framework for interpreting the output of any machine learning

Algorithm 1 Training of the Proposed Model with 80:10:10 Split

```
Require: $D$                                                        ▷ Dataset
Ensure: $M$                                                   ▷ Trained Model
Read data $D$ from the CSV file diabetesmerge.csv
$L \leftarrow D['O']$
$E_Y \leftarrow \text{LE}().fit(L).transform(L)$
$D_Y \leftarrow \text{to\_cat}(E_Y)$
$D \leftarrow D.drop('O', \text{axis} = 1)$
$F_{\text{train}}, F_{\text{val\_test}}, L_{\text{train}}, L_{\text{val\_test}} \leftarrow \text{TRAIN\_TEST\_SPLIT}(D, D_Y, \text{test\_size} = 0.2, \text{shuffle} = \text{True}, \text{random\_state} = 42)$
$F_{\text{val}}, F_{\text{test}}, L_{\text{val}}, L_{\text{test}} \leftarrow \text{TRAIN\_TEST\_SPLIT}(F_{\text{val\_test}}, L_{\text{val\_test}}, \text{test\_size} = 0.5, \text{shuffle} = \text{True}, \text{random\_state} = 42)$
$N \leftarrow \text{Normalization}()$
$N.\text{adapt}(F_{\text{train}})$
$M \leftarrow \text{Sequential}()$
$M.\text{add}(N)$
$M.\text{add}(\text{Dense}(8, \text{activation} =' relu'))$
$M.\text{add}(\text{Dense}(8, \text{activation} =' relu'))$
$M.\text{add}(\text{Dense}(16, \text{activation} =' relu'))$
$M.\text{add}(\text{Dense}(16, \text{activation} =' relu'))$
$M.\text{add}(\text{Dense}(3, \text{activation} =' softmax'))$
$M.\text{compile}(\text{loss} =' categorical\_crossentropy', \text{optimizer} =' RMSprop', \text{metrics} = ['accuracy'])$
$T_C \leftarrow \text{TensorBoard}(\text{log\_dir} = \text{``}logs/f/\text{''} + \text{timestamp})$
$M_C \leftarrow \text{ModelCheckpoint}('output/bestmodel.keras', \text{monitor} =' val\_loss')$
$H \leftarrow M.\text{fit}(F_{\text{train}}, L_{\text{train}}, \text{validation\_data} = (F_{\text{val}}, L_{\text{val}}), \text{verbose} = 1, \text{batch\_size} = 5, \text{epochs} = 100, \text{callbacks} = [T_C, M_C])$
```

model, assigning importance to each feature in the prediction process [58]. They are based on Shapley values from cooperative game theory and provide a principled approach to attributing "credit" among input features.

$$\phi_i = \frac{1}{K} \sum_{k=1}^{K} \left[f(x_k) - E(f(x_k)) \right] \cdot \phi_i^{(k)} \tag{12}$$

In Eq. (12), ϕ_i denotes the SHAP value for feature i, K is the number of samples, $f(x_k)$ represents the model's prediction for sample k, $E(f(x_k))$ is the expected prediction for sample k, and $\phi_i^{(k)}$ indicates the contribution of feature i to sample k. In the DeepNetX2 model, SHAP values are used to explain the importance of features at the global level for the whole dataset and to determine which feature has a major impact on model predictions.

These XAI techniques are useful for explaining and justifying the output of complex models such as DeepNetX2, which in turn would improve their adoptions in real-world use. They are essential for understanding what a specific machine learning model is thinking and why it has made such decisions in the process.

4. Experimental results

In this section, we present our experimental results and discuss the proposed model. The model was assembled on Kaggle's free tier, which offers complementary access to GPU and TPU resources that are essential for training machine-learning models. We utilized mainstream machine-learning libraries, including TensorFlow, Keras, and Scikit-learn, and implemented them in Python as the primary programming language.

4.1. Result analysis

The accuracy, loss, ROC curve, and confusion matrix of machine learning models are pivotal visualizations for understanding and improving the model performance. The accuracy curve illustrates the model's efficacy in accurately predicting outcomes relative to the actual values, with an upward trend indicating improved performance. The loss curve depicts the effectiveness of the model in minimizes errors during training. Training and validation accuracy and loss plots for three different datasets (Type-2 diabetes dataset, local private dataset, and PIMA) are displayed in Fig. 8.

The training and validation accuracy curves revealed differences in model performance across the three datasets over 100 epochs. For the Type-2 Diabetes dataset (see Fig. 8(a)), the large gap initially between the training and validation accuracy suggests the model overfits the training data and may struggle to generalize. The model starts with a low accuracy of approximately 0.5 but rapidly improves, reaching around 0.9 accuracy within the first 10 epochs. Throughout the training process, both the training and validation accuracy consistently remain high, averaging around 0.95–0.98 percent by the end of the 100 epochs. This close alignment between the two curves suggests that the model is well-tuned and effectively generalizes to unseen data. In contrast, the local private dataset (see Fig. 8(b)) shows a close alignment training and validation accuracy curves, indicating that the model can effectively learn and generalize to the validation data. The PIMA Indian dataset (see Fig. 8(c)) exhibits a pattern similar to the Type-2 Diabetes dataset, with the model potentially overfitting. Overall, the local private dataset is the most favorable for the model. The training accuracy starts at approximately 0.5 and increases rapidly, reaching around 0.9 by epoch 20 and continuing to improve slightly until it stabilizes around 0.98 by epoch 60. The validation accuracy also improves steadily, beginning near 0.6, reaching approximately 0.85 by epoch 20, and stabilizing around 0.9 after epoch 40. The Type-2 Diabetes and PIMA Indian datasets present more significant challenges regarding the model's ability to learn and generalize from the data. For the Type-2 Diabetes Dataset (see Fig. 8(d)), started at approximately 0.9 and steadily decreased, flattening out around 0.1 by the end of the 100 epochs, indicating that the model was learning effectively and converging. The validation loss follows a similar trend to the training loss, starting near 0.8 and also decrease to approximately 0.15, stabilizing with minimal fluctuations. This indicates good generalization and minimal overfitting. For the local private dataset (see Fig. 8(e)), the training loss showed a steady decrease, suggesting effective learning by the model. Initially, it decreases but starts to increase around epoch 40, indicating potential overfitting. The model may be memorized in the training data rather than generalize well to new data. For the PIMA Indian Dataset (see Fig. 8(f)), similar to the other datasets, the training loss starts at approximately 0.85 and decreases steadily to approximately 0.1 by 100 epochs, indicating effective learning. The validation loss begins at approximately 0.75, decreases to approximately 0.2, and shows some fluctuations, suggesting good generalization with slight overfitting. However, the Type-2 diabetes dataset demonstrates the most balanced performance with practical learning and generalization. The local private dataset showed signs of overfitting, and the PIMA Indian dataset performs well with minor overfitting.

In Figs. 9(a), 9(b), 9(c), the ROC curve for three different datasets (Type-2 diabetes dataset, local private dataset, and PIMA) are displayed. This curve represents the trade-off between correctly identifying positive instances, and incorrectly classifying negative instances as the classification threshold changes. The area under the ROC curve (AUC) is a pivotal metric for evaluating the performance of a classifier. A higher AUC value signifies a classifier's ability to distinguish between positive and negative instances. As shown in Fig. 9, the DeepNetX2 model exhibits an impressive AUC of 97% on type-2 diabetes dataset, 94% on PIMA Indian dataset, and 90.19% on private dataset. This high AUC confirms that our model's classifier excellently manages to identify positive instances accurately while maintaining a minimum number of false positives. Therefore, our classifier demonstrated strong discrimination capabilities in diabetes diagnosis, differentiating between the positive and negative classes.

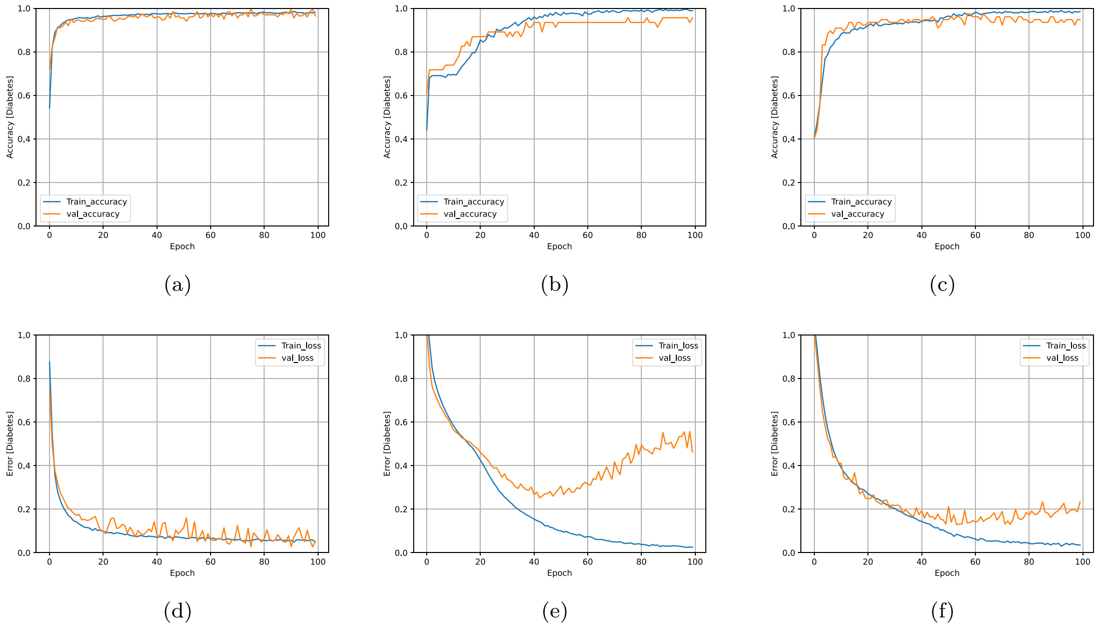


Fig. 8. Evaluation of the DeepNetX2 model's performance across accuracy curves ((a) Type-2 diabetes dataset, (b) Local private dataset, (c) PIMA dataset), and loss curves ((d) Type-2 diabetes dataset, (e) Local private dataset, (f) PIMA dataset for diabetes prediction).

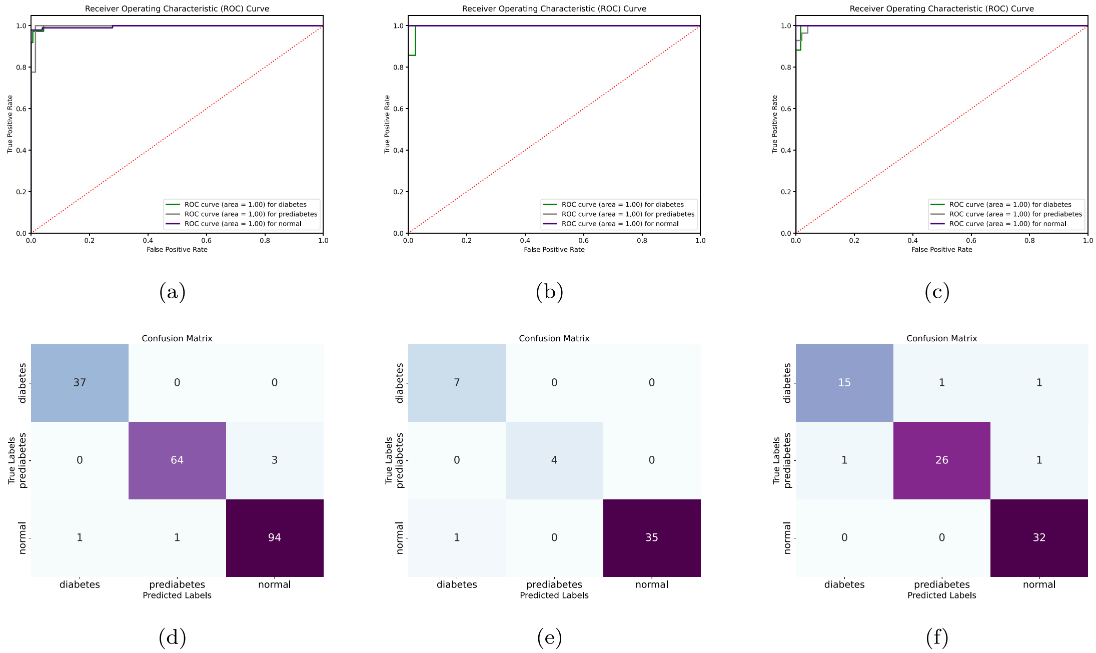


Fig. 9. Evaluation of the DeepNetX2 model's performance across ROC curves ((a) Type-2 diabetes dataset, (b) Local private dataset, (c) PIMA dataset) and Confusion matrices ((d) Type-2 diabetes dataset, (e) Local private dataset, (f) PIMA dataset) for diabetes prediction.

The confusion matrix provides a comprehensive assessment of the performance of a classification model by comparing predicted labels with actual labels. It details model inaccuracies, such as false positives and false negatives and correctly identifies true positives and negatives, thereby providing crucial insights into the precision and efficacy of the model. The confusion matrices of this study are shown in Figs. 9(d), 9(e), and 9(f), for three different datasets (type-2 diabetes dataset, local private dataset, and PIMA). For the Type-2 diabetes dataset (see Fig. 9(d)), the model performed well. It correctly identified all 37 cases of diabetes, 64 of the 67 cases of prediabetes, and 94 of the 96 cases of regular diabetes. There were minimal misclassifications. The model performed well in the local private dataset (see Fig. 9(e)), but with a few more misclassifications of 17 diabetes instances, 15 were correctly classified, with one misclassified as prediabetes and another as usual. Prediabetes instances showed vital accuracy, with 26 out of 28 correctly classified, but one instance was misclassified as diabetes and another as usual. All 32 regular instances were correctly classified. The PIMA Indian dataset (see Fig. 9(f)), demonstrated high accuracy, particularly in prediabetes classification, where all four instances were correctly identified. Out of 8 diabetes instances, seven were correctly classified, with one misclassified as usual. For the standard category, 35 out of 36 instances were correctly classified, with only one misclassified as diabetes. Comparatively, the Type-2 diabetes dataset showed the highest accuracy, especially in diabetes and standard categories, with minimal errors in prediabetes classification. The local private dataset and the PIMA Indian dataset also showed high accuracy, but

Table 8
Performance comparison of Proposed model with and without feature selection (FS).

Set	Model	Dataset	Train accuracy (%)	Test accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
	Without feature selection						
1	ANN	PIMA	85.00	80.50	82.00	80.50	81.20
1	ANN	Local Private	88.00	84.75	85.00	84.75	84.90
1	ANN	Type-2 diabetics	87.50	82.00	83.50	82.00	82.75
1	Gradient Boosting	PIMA	95.00	93.50	94.00	93.50	93.75
1	Gradient Boosting	Local Private	97.00	95.00	96.00	95.00	95.50
1	Gradient Boosting	Type-2 diabetics	96.50	95.00	96.00	95.00	95.50
1	Logistic Regression	PIMA	95.50	90.00	93.00	90.00	91.50
1	Logistic Regression	Local Private	94.00	92.50	91.50	92.50	61.50
1	Logistic Regression	Type-2 diabetics	99.00	90.50	91.00	91.50	96.00
1	DeepNetX2	PIMA	98.37	92.21	94.0	92.0	92.0
1	DeepNetX2	Local Private	99.46	95.74	96.0	96.0	96.0
1	DeepNetX2	Type-2 diabetics	**99.81**	**99.50**	**100**	**99.0**	**100**
With feature selection							
2	ANN	PIMA	63.47	53.24	57.41	53.25	53.58
2	ANN	Local Private	70.25	69.89	79.06	69.89	58.42
2	ANN	Type-2 diabetics	65.24	61.50	66.73	61.50	60.71
2	Gradient Boosting	PIMA	100	100	100	100	100
2	Gradient Boosting	Local Private	100	100	100	100	100
2	Gradient Boosting	Type-2 diabetics	100	100	100	100	100
2	Logistic regression	PIMA	86.80	96.10	96.11	96.10	96.07
2	Logistic regression	Local Private	62.63	68.08	57.0	68.08	62.05
2	Logistic regression	Type-2 diabetics	90.12	91.0	91.24	91.0	91.0
2	DeepNetX2	PIMA	98.21	94.81	95.0	95.0	95.0
2	DeepNetX2	Local Private	**99.19**	**97.87**	**98.0**	**98.0**	**98.0**
2	DeepNetX2	Type-2 diabetics	97.94	97.50	98.0	97.0	97.0

with slightly higher misclassification rates, particularly in the diabetes category. The Type-2 diabetes dataset might contribute to its higher accuracy. In contrast, the local private and PIMA Indian datasets had fewer instances, possibly leading to slightly more errors. Overall, all three models performed well, with the Type-2 diabetes dataset leading to classification accuracy.

4.2. Performance analysis & Model explanations

We employed various performance evaluation metrics to assess the performance of the proposed approach, including the accuracy, precision, recall. Accuracy quantifies the overall correctness of a model by comparing the number of correct predictions with the total number of predictions made. Precision evaluates the model's accuracy for optimistic predictions, recall assesses its ability to identify all true positives, and the f1-score harmonizes precision and recalls into a comprehensive performance metric, which is described mathematically in Eq. (13), (14), (15) and (16).

$$\text{Precision} = \frac{\sum_{i=1}^{n} \text{TP}_i}{\sum_{i=1}^{n} (\text{TP}_i + \text{FP}_i)} \tag{13}$$

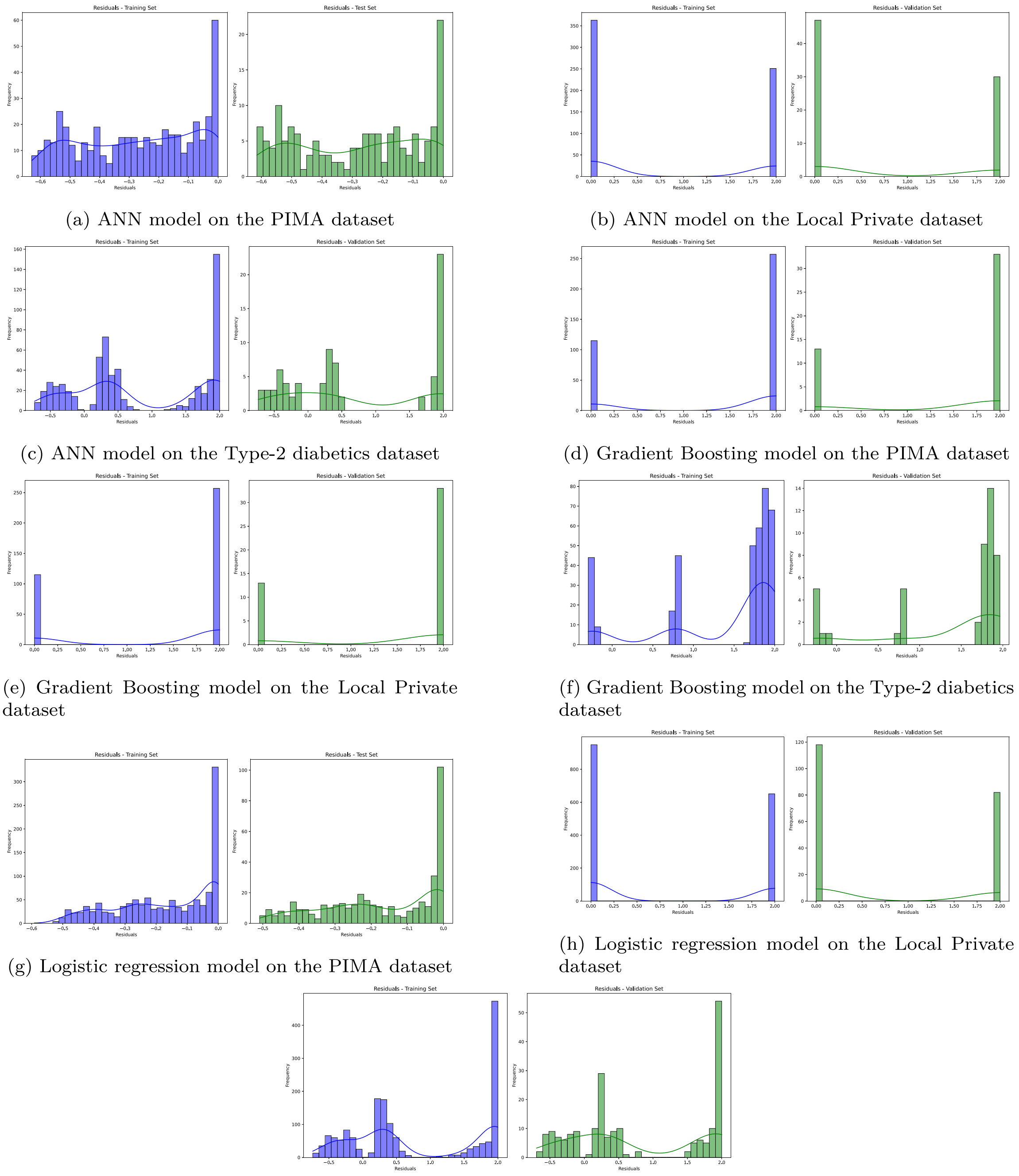
$$\text{Recall} = \frac{\sum_{i=1}^{n} \text{TP}_i}{\sum_{i=1}^{n} (\text{TP}_i + \text{FN}_i)} \tag{14}$$

$$\text{F}_1 = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{15}$$

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} (\text{TP}_i + \text{TN}_i)}{\sum_{i=1}^{n} (\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i)} \tag{16}$$

The following results, shown in Table 8 confirm the improvement of DeepNetX2 model's performance when feature selection (FS) is applied. Comparing DeepNetX2 model with FS and DeepNetX2 model without it, particularly noticeable in the Local Private dataset. When FS was not applied, the model had high training accuracies in all data sets while it produced variable results in test accuracy, although the Local Private data set demonstrated an enhanced measure of precision and recall. Despite the slight decline in the learning process observed for test data such as PIMA with FS testing accuracy, FS further elevates the accuracy value of the Local Private data set (97.87%), indicating better generalization and stability. This improvement emphasizes on the importance of FS in increasing the comprehensibility of the model and its possibility to focus on significant features with the intent of achieving optimal prediction outcomes in analyzed data, especially in the context of healthcare diagnostics. On the other hand, Fig. 10 shows the performance insights of various models shown in Table 8 across pima, type-2 diabetes, and a private dataset. These graphs from the training and validation sets were compared to obtain an overview of the model's effectiveness. As indicated in Table 8, where all metrics are 100% the residuals for the ANN model on the Pima dataset Figs. 10(a), 10(b) and 10(c), these residual plots are distributed throughout a broad range with moderate concentration. In particular, overfitting was evident in the training set, with the model appearing to capture noise. The blocks in the following gradient-boosting image are more equally distributed, indicating that the model is not overfitting, but that the model's learning is not up to par. The blocks indicate that the model is too simple for the dataset and that overfitting of the logistic regression results in large dispersion. The model becomes underfit as a result of this. Similarly, in the rest of Fig. 10, we also obtained almost the same results for the type-2 diabetes dataset and private dataset, suggesting that these models are not appropriate for the datasets.

The proposed DeepNetX2 model excelled in diabetes prediction with statistical robustness, achieving an accuracy of 97.87%, precision and recall of 98.0%, and an F1-score of 98.0%, alongside an AUC of 0.97 as shown in Table 9. It uses LIME and SHAP for explainability which gives easily understandable and intuitive interpretable results, which is vital for clinical uses. Compared to other models such as ANN, SVM, RF, and ensemble methods, which also employ XAI techniques like SHAP, Lime, and Grad-CAM, DeepNetX2's combination of LIME and SHAP offers distinct advantages. LIME has the unique benefit of explaining individual predictions from complex models through local linearization. In contrast, SHAP has developed a unified measure of feature importance and prediction explanations based on game theory.



(a) ANN model on the PIMA dataset

(b) ANN model on the Local Private dataset

(c) ANN model on the Type-2 diabetics dataset

(d) Gradient Boosting model on the PIMA dataset

(e) Gradient Boosting model on the Local Private dataset

(f) Gradient Boosting model on the Type-2 diabetics dataset

(g) Logistic regression model on the PIMA dataset

(h) Logistic regression model on the Local Private dataset

(i) Logistic regression model on the Type-2 diabetics dataset

Fig. 10. Residual plots for different models applied to multiple datasets. The models used are Artificial Neural Networks (ANN), Gradient Boosting, and Logistic Regression. Left side bars represent residuals from the training set, while right side bars represent residuals from the validation set.

4.3. *Using LIME to interpret DeepNetX2*

The prediction probabilities of the classification model are clarified using LIME, as illustrated in Fig. 11. Each consecutive line of the output provides the detailed information on how the model ends with the prediction. "Prediction probabilities" indicate the likelihoods assigned to each possible outcome, with "NOT 1" referring to predictions not belonging to class label 1. The class 0 represents diabetes, class 1 for prediabetes, and class 2 for normal health condition. Additional lines like "Blood Pressure > 80. 00" and "Age ≤ 24. 00" further provide additional conditions or characteristics that define the classes, affecting the model's decisions in regard to each class label chosen. In Fig. 11, the model gives 80% prediction probability of belonging to class 1 which is prediabetes, and just 20% of class 2 which is normal, then it suggests that the individual shows traits in relation to blood pressure or age that are associated with prediabetes. These are obtained by perturbing the input instance slightly and estimating how the model changes locally by deploying an interpretable model. This process significantly enhances the classification model's transparency and interpretability and offers a

Table 9
Comparison of the proposed system with similar recent diabetes prediction works.

Ref.	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	XAI
El Massari et al.	SVM	81.3	81.2	90.4	85.6	0.72	×
Tasin et al.	XGBoost-ADASYN	82	80	81	88.5	0.84	✓
Dutta et al.	Ensemble (DT, RF, XGB, LGB)	83.20	79.65	–	–	0.83	×
Lu et al.	RF, LR, KNN, SVM, NB, DT, XGBOOST, ANN	84.95	85.97	84.95	84.79	0.79 to 0.91	×
Islam et al.	ANN, SVM, RF, XGBoost	90.01	91.80	97.91	94.86	0.85	✓
El-Rashidy et al.	DNN	95.7	95	89	94	0.94	✓
Dharmarathne et al.	XGBoost	77	60	73	65	0.82	✓
Khanna et al.	Stack-Ensemble	96	99	95	–	–	✓
Lalithadevi and Krishnaveni.	OptiDex (DCNN+ECSO)	97.65	93.45	96.46	–	–	✓
Proposed	DeepNetX2	**97.87**	**98.0**	**98.0**	**98.0**	**0.97**	✓

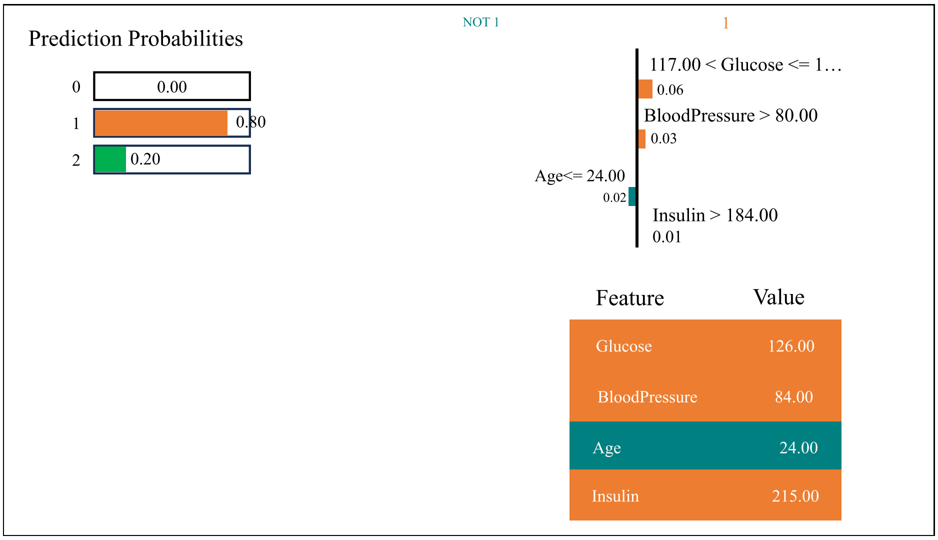


Fig. 11. LIME tabular explainer plots for interpreting individual predictions of the proposed model.

clearer understanding of how it evaluates whether a person may be at risk for prediabetes.

4.4. Using SHAP to interpret DeepNetX2

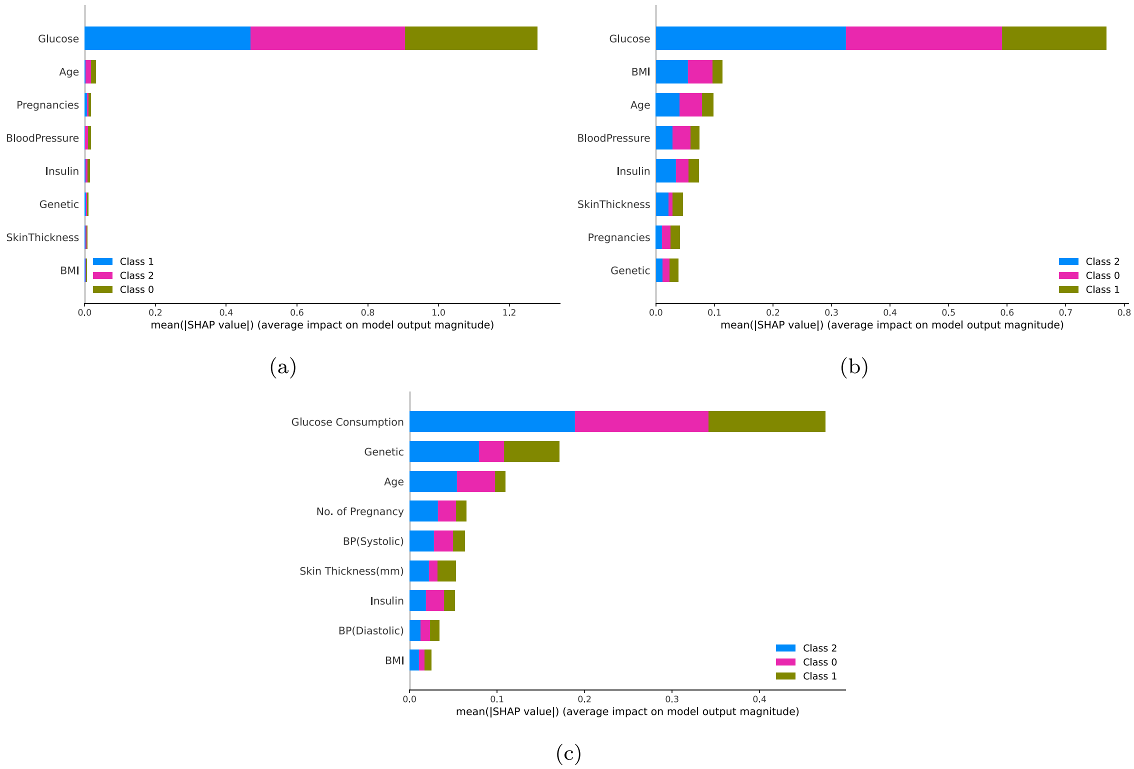
According to Fig. 12, the feature importance for glucose is the highest in all datasets, which means that glucose plays an important role in evaluating the advice of the model among all classes. When the SHAP values for glucose are less than zero, then effectively, it has a negative relation to diabetes and when they are higher than zero, the relationship is positive. As for the other features, they are considered of relatively lesser significance. From these findings, four important characteristics that must be maintained across all datasets were chosen. As accurately demonstrated in Fig. 12, each of the three classes has a different importance level of each feature. This analysis provided the global average SHAP values for the model proposed above.

Subsequently, we conducted a thorough assessment of the key features and analyzed how their values affected the likelihood of diabetes diagnosis. Fig. 13 demonstrates that as the glucose levels increased to 80, the probability of diabetes is not influenced greatly. It has been observed that moderate elevation of glucose levels up to 100 even decrease the chances of a person having diabetes. Similarly, BMI > 30 and age > 30 mainly significantly enhances the probability of diabetes. However, when the glucose concentration is above one hundred, this helps to start adding to the chances of diabetes. In particular, the right side of the figure shows bars of secondary characters that are generally attributed to the primary characteristic. The key factor associated with glucose levels is age, and the primary factor linked with age is glucose. The feature values, indicated by red or blue colors, show heterogeneous variation in diagnosing diabetes. However, there is a notable disparity in glucose levels based on age. Instances where individuals are above 40 years old are marked by a greater occurrence of red dots, signifying that higher glucose levels are more frequently linked to older individuals in the context of diagnosing diabetes.

4.5. Discussion

The proposed model demonstrated remarkable performance in predicting diabetes. It achieved a high accuracy of 97.87% and an impressive F1-score of 98.0%. This level of accuracy outperforms existing models such as Stack-Ensemble (96%) and Lalithadevi and Krishnaveni's OptiDex (97.65%). This highlights DeepNetX2's advanced ability to accurately detect the risk of diabetes. This enhanced accuracy is especially beneficial in clinical settings as it greatly reduces both false positives and false negatives by improving diagnostic precision and patient care. The use of LIME and SHAP with DeepNetX2 added important interpretability to the model. LIME helps provide clear explanations for individual predictions through simpler and more understandable models. SHAP provides a broader view of the most important features, showing that glucose levels play a key role. LIME and SHAP together not only make the model's predictions clearer but also provide detailed insights into how different features interact. This leads to better diagnosis and treatment decisions.

The high diabetes prediction accuracy and robust interpretability of the model have substantial implications for clinical practice. DeepNetX2's clear actionable insights aid healthcare professionals in making informed decisions and personalizing patient care. For instance, the model's emphasis on glucose levels allows clinicians to focus on critical



(a)

(b)

(c)

Fig. 12. Global average interpretation of the proposed model for (a) Type-2 diabetic, (b) PIMA, and (c) Local Private datasets.

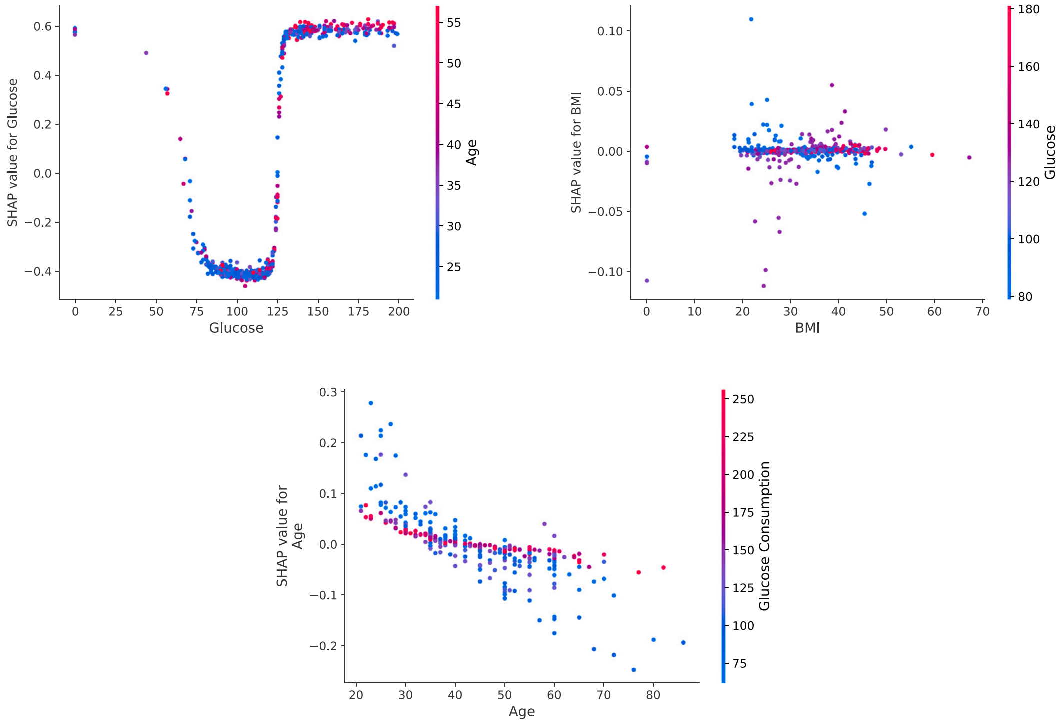


Fig. 13. Feature dependency plots showing the impact of glucose levels and two other features on the proposed model.

risk factors, improving patient outcomes through targeted interventions. Compared to other models employing XAI techniques, such as XGBoost-ADASYN and ensemble methods. DeepNetX2 combination of LIME and SHAP provides a more thorough understanding of prediction dynamics. This dual interpretability framework ensures both local and global perspectives by enhancing the overall transparency and applicability of the model. DeepNetX2 not only excels in predictive accuracy but also offers a transparent, interpretable framework that advances the field of diabetes prediction. Its performance and interpretability make it a leading tool in predictive healthcare analytics.

5. Conclusion and future work

The early detection of diabetes can significantly mitigate the risk of long-term complications. In this study, we present an advanced preprocessing technique along with custom DNN model architecture named DeepNetX2 to precisely predict diabetes in the early stages of a publicly available dataset along with the private dataset. We integrated XAI techniques such as LIME and SHAP with the proposed DeepNetX2 model to make the predictions more transparent, trustworthy, and understandable. Our performance evaluation employed metrics including accuracy, precision, recall, and F1-score, complemented by visualization through an accuracy curve, loss curve, and ROC curve plots. The proposed DeepNetX2 model achieved a remarkable accuracy of 97.87%. Compared to recent related studies, our proposed model showed superior performance. Therefore, our proposed system introduces a robust and reliable approach for predicting and diagnosing diabetes. In future work, we will aim to enhance the security and privacy of our proposed DeepNetX2 architecture for diabetes prediction by integrating it with Decentralized Federated Learning and Blockchain frameworks. This integration is expected to safeguard the sensitive patient data effectively. Additionally, we will focus on real-world applications by developing a smartphone-based application with DeepNetX2's capabilities. This application will facilitate widespread access to advanced diabetes prediction tools by making it easier for users to monitor their health and receive timely interventions. Statistical significance tests will also be performed to further validate the improvements and performance of the model.

Ethical approval

No human and/or animal studies were involved in this study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Sharia Arfin Tanim: Methodology, Formal analysis, Conceptualization. **Al Rafi Aurnob:** Writing – original draft, Methodology, Conceptualization. **Tahmid Enam Shrestha:** Methodology, Formal analysis, Conceptualization. **MD Rokon Islam Emon:** Visualization, Validation, Resources, Methodology. **M.F. Mridha:** Writing – review & editing, Validation, Supervision. **Md Saef Ullah Miah:** Writing – review & editing, Validation, Resources, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. M. F. Mridha reports was provided by American International University Bangladesh. Dr. M. F. Mridha reports a relationship with American International University Bangladesh that includes: employment.

Data availability

This study utilized three datasets:

• The Type-2 Diabetes Dataset can be accessed on the IEEE DataPort: https://ieee-dataport.org/documents/type-2-diabetes-dataset.
• The Local Private Dataset from Pabna Diabetes Hospital, Bangladesh, will be available upon request.
• The PIMA Indian Diabetes Dataset is available on Mendeley: https://data.mendeley.com/datasets/7zcc8v6hvp/1.

Acknowledgment

The authors extend their appreciation to American International University-Bangladesh (AIUB) for funding this research through Researchers Supporting Project Number (AIUB-FST-24-02-01).

References

[1] National Institute of Diabetes, Digestive, K. Diseases, What is diabetes?, 2024, https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes. Online; (Accessed 7 March 2024).

[2] A. Kumar, R. Gangwar, A.A. Zargar, R. Kumar, A. Sharma, Prevalence of diabetes in India: A review of IDF diabetes atlas 10th edition, Curr. Diabetes Rev. 20 (2024) e130423215752, http://dx.doi.org/10.2174/1573399819666230413094200.

[3] National Center for Biotechnology Information, Diabetes mellitus statistics, 2024, https://www.ncbi.nlm.nih.gov/books/NBK581940/#:~:text=An%20estimated%20537%20million%20adults%20aged%2020%E2%80%9379%20years%20worldwide%20to%20be%20living%20with%20diabetes. Online; (Accessed 7 March 2024).

[4] A. Fleck, Who has type 1 diabetes?, 2023, URL https://www.statista.com/statistics/number-of-people-with-type-1-diabetes-worldwide. (Accessed 07 March 2024).

[5] A. Auvinen, K. Luiro, J. Jokelainen, I. Järvelä, M. Knip, J. Auvinen, J. Tapanainen, Type 1 and type 2 diabetes after gestational diabetes: a 23 year cohort study, Diabetologia 63 (2020) 2123–2128.

[6] L. García-Flores, S. Medina, J. Morales-Ortiz, M. Olvera-Soto, L. Velázquez-Villegas, A. Tovar, N. Torres, P. Ortíz-Vilchis, J. Castellanos, A. Aguilar-Martínez, F. Marquez-Sandoval, J. Curiel, L. Valle-Mondragón, Antioxidants in sport and exercise: More than just protecting against exercise-induced oxidative stress, Antioxidants 10 (8) (2021) 1318, http://dx.doi.org/10.3390/antiox10081318.

[7] A. Bovolini, J. Garcia, M. Andrade, J. Duarte, Metabolic syndrome pathophysiology and predisposing factors, Int. J. Sports Med. 42 (03) (2021) 199–214.

[8] M. Li, J.-R. Lan, J.-L. Liang, X.-L. Xiong, Diagnostic accuracy of fasting plasma glucose as a screening test for gestational diabetes mellitus: a systematic review and meta-analysis, Eur. Rev. Med. Pharmacol. Sci. 24 (21) (2020).

[9] F.Y. Kuo, K.-C. Cheng, Y. Li, J.-T. Cheng, Oral glucose tolerance test in diabetes, the old method revisited, World J. Diabetes 12 (6) (2021) 786.

[10] M. Evans, Z. Welsh, A. Seibold, Reductions in HbA1c with flash glucose monitoring are sustained for up to 24 months: a meta-analysis of 75 real-world observational studies, Diabetes Therapy 13 (6) (2022) 1175–1185.

[11] Y.B. Özçelik, A. Altan, Classification of diabetic retinopathy by machine learning algorithm using entorpy-based features, in: Proceedings of the ÇAnkaya International Congress on Scientific Research, IKSAD Golbasi, Adiyaman Province, Turkey, 2023, pp. 10–12.

[12] O. Dweekat, S. Lam, Optimized design of hybrid genetic algorithm with multilayer perceptron to predict patients with diabetes, Soft Comput. 27 (10) (2023) 6205–6222.

[13] S. Alex, J. Nayahi, H. Shine, V. Gopirekha, Deep convolutional neural network for diabetes mellitus prediction, Neural Comput. Appl. 34 (2) (2022) 1319–1327.

[14] C. Olisah, L. Smith, M. Smith, Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective, Comput. Methods Programs Biomed. 220 (2022) 106773.

[15] R. Krishnamoorthi, S. Joshi, H.Z. Almarzouki, P.K. Shukla, A. Rizwan, C. Kalpana, B. Tiwari, [Retracted] a novel diabetes healthcare disease prediction framework using machine learning techniques, J. Healthc. Eng. 2022 (1) (2022) 1684017.

[16] T.R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P.K.R. Maddikunta, G. Srivastava, Deep neural networks to predict diabetic retinopathy, J. Ambient Intell. Humaniz. Comput. (2023) 1–14.

[17] M.F. Aslan, K. Sabanci, A novel proposal for deep learning-based diabetes prediction: Converting clinical data to image data, Diagnostics 13 (4) (2023) 796.

[18] O.S. Zargar, A. Bhagat, T.A. Teli, S. Sheikh, Early prediction of diabetes mellitus on pima dataset using ML and DL techniques, J. Army Eng. Univ. PLA (2023).

[19] Y.B. Özçelik, A. Altan, Overcoming nonlinear dynamics in diabetic retinopathy classification: a robust AI-based model with chaotic swarm intelligence optimization and recurrent long short-term memory, Fractal Fract. 7 (8) (2023) 598.

[20] A. Mashraqi, B. Allehyani, Current trends on the application of artificial intelligence in medical sciences, Bioinformation 18 (11) (2022) 1050.

[21] G. Lima, N. Grgić-Hlača, J.K. Jeong, M. Cha, The conflict between explainable and accountable decision-making algorithms, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 2103–2113.

[22] IBM, What is explainable AI?, 2024, URL https://www.ibm.com/topics/explainable-ai. (Accessed 29 August 2024).

[23] M.A.R. Refat, M. Al Amin, C. Kaushal, M.N. Yeasmin, M.K. Islam, A comparative analysis of early stage diabetes prediction using machine learning and deep learning approach, in: 2021 6th International Conference on Signal Processing, Computing and Control, ISPCC, IEEE, 2021, pp. 654–659.

[24] U. Ahmed, G.F. Issa, M.A. Khan, S. Aftab, M.F. Khan, R.A. Said, M. Ahmad, Prediction of diabetes empowered with fused machine learning, IEEE Access 10 (2022) 8529–8538.

[25] H.F. Ahmad, H. Mukhtar, H. Alaqail, M. Seliaman, A. Alhumam, Investigating health-related features and their impact on the prediction of diabetes using machine learning, Appl. Sci. 11 (3) (2021) 1173.

[26] J. Ramesh, R. Aburukba, A. Sagahyroon, A remote healthcare monitoring framework for diabetes prediction using machine learning, Healthc. Technol. Lett. 8 (3) (2021) 45–57.

[27] H.M. Deberneh, I. Kim, Prediction of type 2 diabetes based on machine learning algorithm, Int. J. Environ. Res. Public Health 18 (6) (2021) 3317.

[28] R.D. Joshi, C.K. Dhakal, Predicting type 2 diabetes using logistic regression and machine learning approaches, Int. J. Environ. Res. Public Health 18 (14) (2021) 7346.

[29] A.Z. Peng, X.H. Kong, S.T. Liu, H.F. Zhang, L.L. Xie, L.J. Ma, Y. Chen, Explainable machine learning for early predicting treatment failure risk among patients with TB-diabetes comorbidity, Sci. Rep. 14 (1) (2024) 6814.

[30] C.-Y. Chou, D.-Y. Hsu, C.-H. Chou, Predicting the onset of diabetes with machine learning methods, J. Pers. Med. 13 (3) (2023) 406.

[31] E. Dritsas, M. Trigka, Data-driven machine-learning methods for diabetes risk prediction, Sensors 22 (14) (2022) 5304.

[32] T.R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P.K.R. Maddikunta, G. Srivastava, Deep neural networks to predict diabetic retinopathy, J. Ambient Intell. Humaniz. Comput. (2023) 1–14.

[33] P. Madan, V. Singh, V. Chaudhari, Y. Albagory, A. Dumka, R. Singh, A. Gehlot, M. Rashid, S. Alshamrani, A. AlGhamdi, An optimization-based diabetes prediction model using CNN and bi-directional LSTM in real-time environment, Appl. Sci. 12 (2022) 3989, http://dx.doi.org/10.3390/app12083989.

[34] A.N. Soni, Diabetes mellitus prediction using ensemble machine learning techniques, 2020, Available at SSRN 3642877.

[35] T.R. Mahesh, D. Kumar, V.V. Kumar, J. Asghar, B.M. Bazezew, R. Natarajan, V. Vivek, Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease, Comput. Intell. Neurosci. (2022).

[36] K. Abnoosian, R. Farnoosh, M.H. Behzadi, Prediction of diabetes disease using an ensemble of machine learning multi-classifier models, BMC Bioinform. 24 (1) (2023) 337.

[37] H. Zhou, Y. Xin, S. Li, A diabetes prediction model based on boruta feature selection and ensemble learning, BMC Bioinform. 24 (1) (2023) 224.

[38] H. Nemat, H. Khadem, M.R. Eissa, J. Elliott, M. Benaissa, Blood glucose level prediction: advanced deep-ensemble learning approach, IEEE J. Biomed. Health Inf. 26 (6) (2022) 2758–2769.

[39] H. Kibria, M. Nahiduzzaman, M. Goni, M. Ahsan, J. Haider, An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI, Sensors 22 (19) (2022) 7268, http://dx.doi.org/10.3390/s22197268.

[40] M. Obayya, N. Nemri, M.K. Nour, M. Al Duhayyim, H. Mohsen, M. Rizwanullah, A. Motwakel, Explainable artificial intelligence enabled TeleOphthalmology for diabetic retinopathy grading and classification, Appl. Sci. 12 (17) (2022) 8749.

[41] K. Sękowski, J. Grudziąż-Sękowska, J. Pinkas, M. Jankowski, Public knowledge and awareness of diabetes mellitus, its risk factors, complications, and prevention methods among adults in Poland-A 2022 nationwide cross-sectional survey, Front. Public Health 10 (2022) 1029358, http://dx.doi.org/10.3389/fpubh.2022.1029358.

[42] H. Lu, S. Uddin, F. Hajati, et al., A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus, Appl. Intell. 52 (2022) 2411–2422.

[43] A. Dutta, M.K. Hasan, M. Ahmad, M.A. Awal, M.A. Islam, M. Masud, H. Meshref, Early prediction of diabetes using an ensemble of machine learning models, Int. J. Environ. Res. Public Health 19 (19) (2022) 12378.

[44] H. El Massari, Z. Sabouri, S. Mhammedi, N. Gherabi, Diabetes prediction using machine learning algorithms and ontology, J. ICT Stand. 10 (2) (2022) 319–337.

[45] I. Tasin, T.U. Nabil, S. Islam, R. Khan, Diabetes prediction using machine learning and explainable AI techniques, Healthc. Technol. Lett. 10 (1–2) (2023) 1–10.

[46] M.M. Islam, M.J. Rahman, M.S. Rabby, M.J. Alam, S.A.I. Pollob, N.F. Ahmed, M. Maniruzzaman, Predicting the risk of diabetic retinopathy using explainable machine learning algorithms, Diabetes Metab. Syndr.: Clin. Res. Rev. 17 (12) (2023) 102919.

[47] N. El-Rashidy, N.E. ElSayed, A. El-Ghamry, F.M. Talaat, Utilizing fog computing and explainable deep learning techniques for gestational diabetes prediction, Neural Comput. Appl. 35 (10) (2023) 7423–7442.

[48] B. Lalithadevi, S. Krishnaveni, Diabetic retinopathy detection and severity classification using optimized deep learning with explainable AI technique, Multimedia Tools Appl. (2024) 1–65.

[49] G. Dharmarathne, T.N. Jayasinghe, M. Bogahawaththa, D.P.P. Meddage, U. Rathnayake, A novel machine learning approach for diagnosing diabetes with a self-explainable interface, Healthc. Anal. 5 (2024) 100301.

[50] V. Vivek Khanna, K. Chadaga, N. Sampathila, S. Prabhu, P.R. Chadaga, D. Bhat, S. KS, Explainable artificial intelligence-driven gestational diabetes mellitus prediction using clinical and laboratory markers, Cogent Eng. 11 (1) (2024) 2330266.

[51] A.D. Association, Standards of medical care in diabetes—2014, Diabetes Care 37 (Supplement_1) (2014) S14–S80.

[52] D. Shang, A. Li, P. Shang, An improved nonlinear correlation method for feature selection of complex data, Nonlinear Dynam. 111 (2023) 11357–11369, http://dx.doi.org/10.1007/s11071-023-08406-w.

[53] Google Developers, Deep neural network models for recommendation, 2024, https://developers.google.com/machine-learning/recommendation/dnn/softmax. (Accessed 18 March 2024).

[54] Z.Y. Khan, Z. Niu, CNN with depthwise separable convolutions and combined kernels for rating prediction, Expert Syst. Appl. 170 (2021) 114528.

[55] S. Hu, Z. Lou, X. Yan, Y. Ye, A survey on information bottleneck, IEEE Trans. Pattern Anal. Mach. Intell. (2024).

[56] Wikipedia contributors, Explainable artificial intelligence, 2024, https://en.wikipedia.org/wiki/Explainable_artificial_intelligence. (Accessed 18 March 2024).

[57] J. Shin, Feasibility of local interpretable model-agnostic explanations (LIME) algorithm as an effective and interpretable feature selection method: comparative fNIRS study, Biomed. Eng. Lett. 13 (2023) 689–703, http://dx.doi.org/10.1007/s13534-023-00291-x.

[58] K. Roshan, A. Zafar, Utilizing XAI technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation (SHAP), 2021, arXiv preprint arXiv:2112.08442.