Journal of Posthumanism

2025
Volume: 5, No: 12, pp. 83–98
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i12.3723

Early Prognosis of Diabetes Harnessing Physiological Data and Artificial Intelligence

Ayman Sabih[1], Rumana Islam[1], Mohammed Tarique[1]

Abstract

This paper presents an artificial intelligence-based early detection technique for individuals with diabetes. According to the International Diabetes Federation (2024), the number of adults with diabetes is projected to exceed 850 million by 2050. Unhealthy food habits, physical inactivity, and family history are commonly blamed for diabetes. Improperly managed diabetes can lead to life-threatening complications, including cardiovascular diseases, kidney failure, nerve damage, and vision problems. Hence, early detection of diabetes is crucial and has become a primary focus of recent research. Computer-based disease detection, powered by artificial intelligence, can play a pivotal role here. With recent advances in algorithms and artificial intelligence, these technologies have become increasingly popular across diverse fields of biomedicine and bioinformatics, leading to rapid advancements in computer-based disease diagnosis. This work investigates the Pima Indian Diabetes Dataset and demonstrates that a shallow feedforward neural network (FFNN) can predict diabetes from critical biological data, achieving 79.1% accuracy. This population- based projection measure can effectively alert individuals to be vigilant and participate in recommended health screenings.

Keywords: *Artificial intelligence, Classification, Data analysis, Diabetes, Neural network, Performance, Prediction.*

Introduction

Diabetes is a common chronic disease among the world's population. Currently, almost 590 million people have diabetes, and this figure is projected to reach 850 million by the year 2050 (International Diabetes Federation, 2025). Diabetes also caused 3.4 million deaths in 2024. Eighty percent (80%) of the diabetic patients belong to low and middle-income countries (World Health Organization, 2025), overwhelming their country's healthcare systems. Health expenditure related to diabetes is skyrocketing and has already crossed 1.015 trillion US dollars: a 338% increase over the last 17 years. The most alarming issue is that 252 million people worldwide are still living with undiagnosed diabetes and are exposed to serious health hazards. Diabetes is directly related to insulin, a hormone that regulates blood glucose levels. Diabetes occurs mainly for two reasons: (a) the pancreas does not produce enough insulin in the body, and (b) the human body cannot effectively use the insulin.

There are three types of diabetes: Type 1, Type 2, and gestational diabetes. Type 1 diabetes typically begins in childhood or adolescence and is characterized by a deficiency of insulin-producing cells in the pancreas. Hence, our body produces little insulin to regulate the sugar level.

[1] Department of Electrical and Computer Engineering, Fujairah University, United Arab Emirates (UAE), E-mail: m.tarique@fu.ac.ae

Occasionally, Type 1 diabetes can also occur in adulthood. The lifelong insulin therapy is the sole medical option for these patients. Type 2 diabetes develops in adults and is caused by insulin resistance. Increasingly, children are being diagnosed with this type of diabetes. Unhealthy lifestyle, genetics, and obesity are the contributing factors for this type of diabetes. Gestational diabetes often occurs during pregnancy when a mother undergoes hormonal changes, impairing insulin production in her body. Also, women experience hormonal changes during their menstrual cycle and menopause. These hormonal changes negatively affect insulin action, making it difficult to predict and manage blood glucose levels. Some of the symptoms experienced by women are like those experienced by men. Still, some symptoms are experienced only by women, for example: prone to urinary infections, polycystic ovarian syndrome, female sexual dysfunction, vaginal and oral yeast infections, including vaginal thrush.

The common symptoms of diabetes include increased thirst, frequent urination, unexplained weight loss, increased hunger, fatigue, blurred vision, slow-healing sores, and numbness in the feet and arms. If untreated, increased sugar levels (also called hyperglycemia) can cause other life-threatening diseases, including nerve damage, kidney failure, Alzheimer’s disease, depression, and cardiovascular diseases. In general, the pathophysiology of diabetes mellitus is characterized by hyperinsulinemia. That in turn causes progressive loss of the function of $\boldsymbol{\beta}$-cells, resulting in this disease (Figure 1).

Till now, this disease is not curable, but manageable with prescribed insulin and other medicines. Specifically, numerous lives have been saved since the discovery of insulin at the University of Toronto (Best C. H., & Scott, D. A., 1923; de Leiva-Hidalgo A., & de Leiva-Pe´rez A., 2023). To date, insulin is commonly prescribed by endocrinologists as a life-saving antidiabetic medication (Abel, J.J., 1926; Crowfoot, D., 1935; Sanger, F., 1960; Vecchio, I. et al., 2018). Nevertheless, early detection of diabetes is crucial for effective disease management, allowing patients to live longer and healthier lives.

Recently, computer-based diagnosis has been playing an increasingly important role in biomedicine and bioinformatics. The main credit goes to the recent developments in artificial intelligence algorithms. Disease prediction using artificial intelligence (AI) and machine learning (ML) is one of the fastest-emerging applications in recent years. The AI-based diagnosis has been effectively adopted to detect diseases, including breast cancer, skin cancer, lung cancer, leukemia, heart disease, chronic kidney disease, liver disease, COVID-19, and, not least, diabetes (Malakar, S. et al., 2022; Islam, R., & Tarique, M., 2022; Islam, R., & Tarique, M., 2024). Highly motivated researchers from computer science, biology, medicine, statistics, and drug discovery are working relentlessly to develop new algorithms and systems that achieve near-perfect results in detecting diabetes.

Related Works

A considerable amount of work has been published on computer-based detection of diabetes. Some of these recent, directly related to this work, are worth mentioning here.

In (Tasnin, I. et al. 2022), the authors have employed decision trees, support vector machines (SVMs), random forests, logistic regression, k-nearest neighbors (KNNs), and various ensemble techniques to detect diabetic patients. They applied the synthetic minority over-sampling technique (SMOTE) and adaptive synthetic sampling (ADSYN) to deal with imbalanced data. The authors achieved the highest accuracy of 81% using the ensemble classifier (XGBoost) and ADASYN data balancing. They applied local interpretable model-agnostic explanation (LIME) and Shapley additive explanations (SHAP) techniques to explain the prediction results.

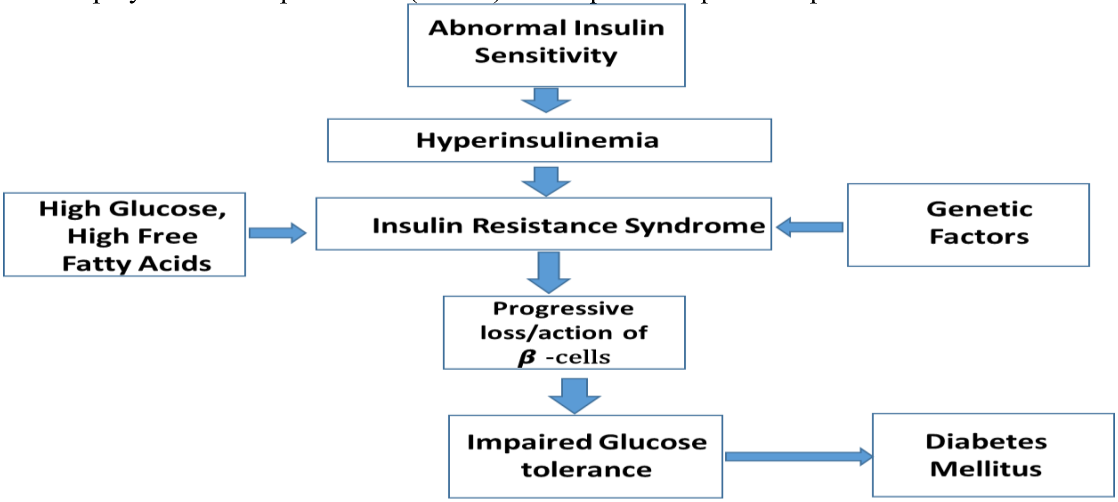

Abnormal Insulin Sensitivity
Hyperinsulinemia
High Glucose, High Free Fatty Acids
Insulin Resistance Syndrome
Genetic Factors
Progressive loss/action of β -cells
Impaired Glucose tolerance
Diabetes Mellitus

Figure 1. The pathophysiology of diabetes.

In a similar study (Ahmed, A. et al., 2024), the authors employed random forests, decision trees, Naïve Bayes, and logistic regression to detect diabetes in the female population. They demonstrated that random forests outperform other algorithms. They achieved 80% accuracy, 82% precision, 88% sensitivity, and 20% error rate. The authors also applied principal component analysis (PCA) for dimension reduction in their work.

In (Toleva et al., 2025), the authors utilized random forests and support vector machines (SVMs) to identify diabetic patients. They adopted two methods. In the first method, they used classical classification methods for imbalanced data. In the second method, they used resampling and shuffling to address data imbalance. They demonstrated that the proposed method outperforms the classical method, achieving 95.05% accuracy.

In (Naz, H., & Ahuja, S., 2020), the authors investigated diabetic detection using artificial neural networks (ANNs), Naïve Bayes (NB), decision trees, and deep learning. They achieved the best performance with the deep learning techniques. The authors achieved an accuracy of 98.07%, a precision of 95.22%, a recall of 98.46%, a specificity of 99.29%, and a sensitivity of 95.52%.

A novel e-diagnosis system, based on machine learning and the Internet of Medical Things (IoMT), has been presented in (Chang et. al., 2023) to detect Type 2 diabetes. The authors applied supervised machine learning models, namely the NB classifier, random forests, and decision trees. They investigate three cases: 3-factor, 5-factor, and all features in the work. The 3-factor features include glucose, BMI, and age, and the 5-factor features include glucose, BMI, age, insulin, and skin thickness. In the third case, they include all features listed in the PIMA database. The authors achieved the highest accuracy of 80% with random forests, including all features.

In (Khanam, J.J., & Foo, S. Y., 2021), a performance comparison of various machine learning methods for detecting diabetes is presented. The authors investigated the performance of NB, SVM, linear regression, AdaBoost, random forests, k-nearest neighbors (kNN), decision trees, and artificial neural networks (ANNs) in detecting diabetes. The authors have achieved the highest accuracy of 88.6% with a two-hidden-layer-based ANN.

Two types of networks, a feedforward neural network (FFNN) and a convolutional neural network (CNN), are employed to detect diabetes in (Ashour, A. F., et al., 2024). The FFNN model achieved 82% accuracy, whereas the CNN achieved 80.52%.

A novel standard deviation k-nearest neighbor (SDKNN) algorithm was presented in (Petra, R. & Khunba, B., 2020) to detect diabetes. The authors have investigated the performance of their proposed algorithm with k=5, 6, 7, 10, 15, 20, 25, and 27. The average detection accuracy was 83.2%. Long short-term memory (LSTM), random forest, and CNN have been investigated in (Mousa, A., et al., 2023) for detecting diabetes. The results show that the LSTM achieved the highest accuracy of 85%. On the other hand, random forests and CNNs achieved the highest accuracies of 78% and 82%, respectively.

An ensemble classifier was applied in (Deepalakshmi, M. et al., 2025) to detect diabetes. The authors adopted the firefly algorithm to optimize hyperparameters and achieve the best results. They have undergone a thorough pre-treatment process, including handling missing values, normalization, and outlier detection, in their proposed work. Their ensemble model achieved a remarkable accuracy of 97.27%.

A novel approach for converting numerical data from the PIMA diabetic dataset into visual representations (i.e., images) has been proposed in (Zarger, O. S. et al., 2024). The authors claimed that this type of dataset conversion enables them to employ powerful CNN models, including VGG16 and ResNet50, to achieve very high accuracy (97.19%). This high accuracy suggested that their approach for converting numerical data into an image was very effective for early diabetes diagnosis.

In (Hounge, P. & Bigirimana, A. G., 2022), the authors employed a deep neural network (DNN) to predict diabetes with 10-fold cross-validation and achieved an accuracy of 89%. Three types of classifiers —tree-based, function-based, and rule-based —were investigated in (Huynh, H. V. T., et al., 2024) for diagnosing diabetes. The tree-based classifiers include random forest, J48, and random tree. Among these algorithms, random forest achieved the highest accuracy of 75.8%. The authors also include function-based classifiers, namely log, multi-layer perceptron (MLP), and stochastic gradient descent (SGD), in their investigation. Among these three algorithms, SGD achieved the highest accuracy of 78%. Among the three rule-based classifiers —JRip, OneR, and PART —JRip achieved the highest accuracy of 76%.

A comprehensive investigation on diabetes detection using three datasets (PIMA Indian Database, 2025; Diabetic Dataset 2025; & BIT2019 dataset), presented in (Abousaber, J. et al., 2025). The authors adopted several machine learning algorithms —logistic regression, kNN, decision trees, random forests, gradient boosting, SVM, NB, and XGBoost—in that investigation. To address data imbalance, the authors have applied several data balancing algorithms: ADASYN, SMOTE, and Borderline SMOTE. They achieved 100% accuracy with random forests, XGBoost, and LightGBM, using ADASYN, SMOTE, and Borderline-SMOTE data balancing techniques. They improved the detection accuracy by 9.13%-15.22%, incorporating these data balancing techniques in their work.

In this paper, a feedforward neural network (FFNN) is employed to evaluate the efficacy of a diabetes detection algorithm based on biological features. The main contributions of the proposed algorithm are as follows: (a) it employs eight critical physiological data points of diabetic patients to train an FFNN, (b) it investigates an FFNN's capability for diabetes identification, and (c) it reduces the substantial computational burden compared to other existing works as mentioned above.

Materials and Methods

Database

The PIMA Indians Diabetes dataset is a collection of medical data of diabetic patients of PIMA Indian heritage (PIMA Indian Dataset, 2025). It includes eight features: pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, and age. The binary outcomes are "0" for no diabetes and "1" for diabetes. The dataset is derived from the National Institute of Diabetes and Digestive and Kidney Diseases and is commonly used in machine learning for diabetes prediction tasks. The features are directly related to diabetes as follows.

- Pregnancies: Substantial maternal metabolic and lifestyle alterations during pregnancy lead to diabetes. A study shows that women with three or more pregnancies have a higher chance of diabetes.
- Glucose: Normal blood sugar is 99 mg/dL or lower after a fasting blood test. A blood sugar of 100 mg/dL or higher is considered abnormal. A range of 100-125 mg/dL falls under the category of prediabetes, while a blood sugar of 126 mg/dL or higher is considered Type 2 diabetes (Liu, B. et al., 2020).
- Blood Pressure: Diabetic patients are twice as likely to have high blood pressure because high blood sugar damages blood vessels, narrowing arteries (Gunther, A. et al.,2025).
- Skin Thickness: Triceps skin fold thickness (TSF) (mm) has varying associations with Type 2 diabetes.
- Insulin: Insulin levels, measured by two-hour serum insulin (μU/ml), are directly related to differentiating the diabetic population.
- BMI: Body mass index (BMI) is a tool used by healthcare providers to estimate body fat based on height and weight. It can help assess risk factors for specific health ailments. Although it is not always an accurate representation of body fatness, a high body mass index (BMI) is strongly associated with an increased risk of developing type 2 diabetes in humans. The BMI is measured using the mass index, i.e., weight (*kg*) divided by height (*m*) squared i.e., (kg/m^2).
- Diabetes Pedigree Function (DPF): The DPF scores the likelihood of diabetes based on a person's family history. The DPF has a realistic range of 0.08-2.42.
- Age: Age (in years) contributes to diabetes when the body's cells do not respond effectively to insulin. The pancreas's ability to secrete enough insulin to manage blood sugar levels also reduces with age.

Data Cleansing

The dataset consists of 767 instances, all female patients of Pima Indian heritage aged 21 years or older. Among these data, 267 of the patients have developed diabetes, and the remaining 500 patients are without diabetes. The distribution of the data is illustrated in Figure 2, which shows that 65% of the samples belong to patients without diabetes, and the remaining 35% to patients with diabetes. In the existing database, the authors have noticed some unusual values in the patients' data. For example, skin thickness, body mass index (BMI), and blood pressure cannot be zero. Similarly, there are unusual pregnancies (e.g., 17) recorded in the data. Hence, the authors judiciously replace these data with the average value of the remaining valid data. The statistical analysis of the PIMA database is listed in Table 1. This demonstrates that the features recorded in the PIMA database are indeed essential to distinguish between patients with and without diabetes. For example, the average value of pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, DPF, and age is higher in patients with diabetes compared to those without diabetes. The maximum values of these features are also higher in patients with diabetes than in those without diabetes, as shown in Table 1.

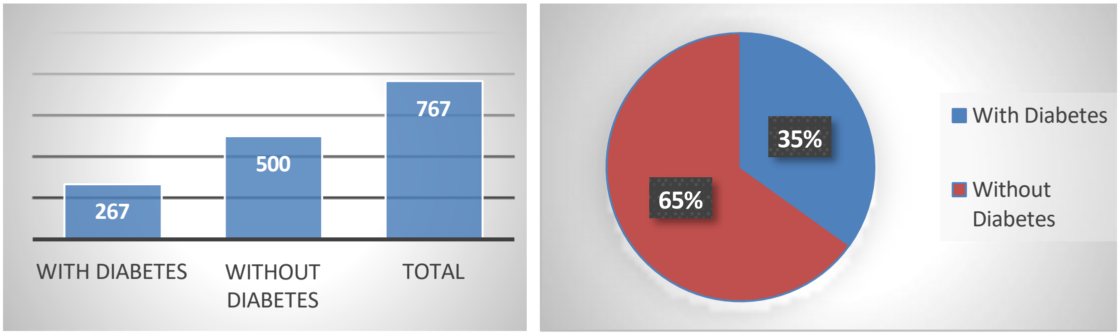


Figure 2. The data distribution in the PIMA database.

The correlation among these features is illustrated in Figure 3. This figure shows that some features are more correlated than others. For example, the correlation between skin thickness and insulin is highly correlated. On the other hand, the DPF and age are lowly correlated. The box plots of the features are shown in Figure 4. This figure shows that the insulin feature has the most outliers among all features. On the other hand, the pregnancies and DPF have the lowest outliers.

Feed Forward Neural Network (FFNN)

A feed-forward neural network (FFNN) has been employed in this work to detect diabetes. In this type of artificial neural network, information flows in a single direction: from the input layer through the hidden layers to the output layer, without loops or feedback, as shown in Figure 5. It is widely used for pattern recognition tasks. The FFNN used in this work has a structured layer as follows.

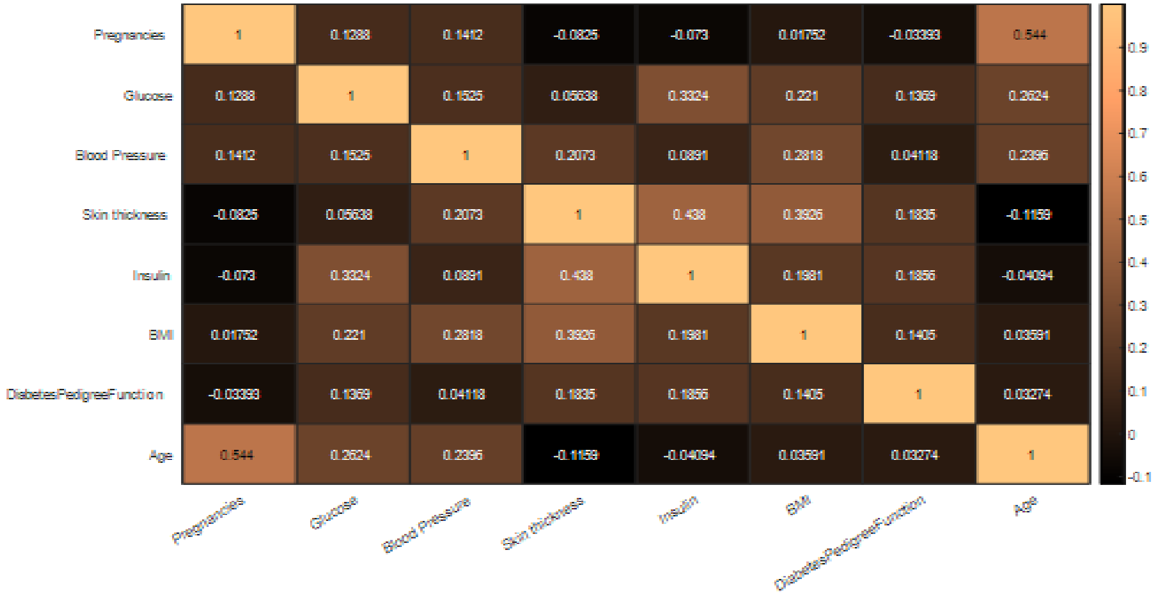
- Input Layer: The input layer consists of neurons that receive the input data. Each neuron in the input layer represents a feature of the input data. In this case, eight neurons for the eight features of each patient are employed in the input layer.
- Hidden Layers: Forty (40) neurons are placed between the input and output layers. This layer is responsible for learning the complex patterns in the data. Each neuron in a hidden

layer applies a weighted sum of inputs followed by a Sigmoid function.

- Output Layer: The output layer provides the network's final output. The number of neurons in this layer corresponds to the number of classes in a classification problem or the number of outputs in a regression problem. In this work, there is only one neuron as the final output decision is binary. The SoftMax activation function has been used at the output layer.

Table 1. The statistical analysis of the investigated features.

Features	**With Diabetics**						**Without Diabetics**				
	Min.	Max.	Avg.	Std. dev.	median	mode	Min.	Max.	Avg.	Std. dev.	med
Pregnancies	0	17	**5**	4	3	1	0	13	**2.72**	2.6	2
Glucose (mg/dl)	78	198	**145**	30	145	128	56	197	**111**	24.64	107.
Blood Pressure	30	110	**74**	13	74	70	24	106	**69**	11.89	70
Skin Thickness(mm)	7	63	**33**	9.6	33	32	7	60	**27**	10.43	27
Insulin	14	846	**207**	132.70	169.5	130	15	744	**131**	102.6	105
BMI(kg/m^2)	22.9	67.1	**36**	6.73	34.6	31.60	18.2	57.3	**32**	6.79	31.2
DPF	0.127	2.42	**0.63**	0.41	0.546	0.254	0.085	2.329	**0.47**	0.299	0.41
Age (years)	21	60	**36**	11	33	25	21	81	**28**	8.90	25



	Pregnancies	Glucose	Blood Pressure	Skin thickness	Insulin	BMI	DiabetesPedigreeFunction	Age
Pregnancies	1	0.1288	0.1412	-0.0825	-0.073	0.01752	-0.03393	0.544
Glucose	0.1288	1	0.1525	0.05638	0.3324	0.221	0.1369	0.2624
Blood Pressure	0.1412	0.1525	1	0.2073	0.0891	0.2818	0.04118	0.2396
Skin thickness	-0.0825	0.05638	0.2073	1	0.438	0.3926	0.1835	-0.1159
Insulin	-0.073	0.3324	0.0891	0.438	1	0.1981	0.1856	-0.04094
BMI	0.01752	0.221	0.2818	0.3926	0.1981	1	0.1405	0.03591
DiabetesPedigreeFunction	-0.03393	0.1369	0.04118	0.1835	0.1856	0.1405	1	0.03274
Age	0.544	0.2624	0.2396	-0.1159	-0.04094	0.03591	0.03274	1

Figure 3. The correlation matrix.

Each connection between neurons in these layers has an associated weight that is adjusted during the training process to minimize prediction errors. The FFNN is chosen for the following reasons: (a) the FFNN has a straightforward architecture, making them easy to design and implement, (b) the one-way data flow assists in efficient computation, making the network suitable for real-time applications and scenarios with limited system resources, (c) it can be applied to a wide range of tasks, including classification, regression, and prediction, and (d) it excels in situations where the input data points are independent of each other.

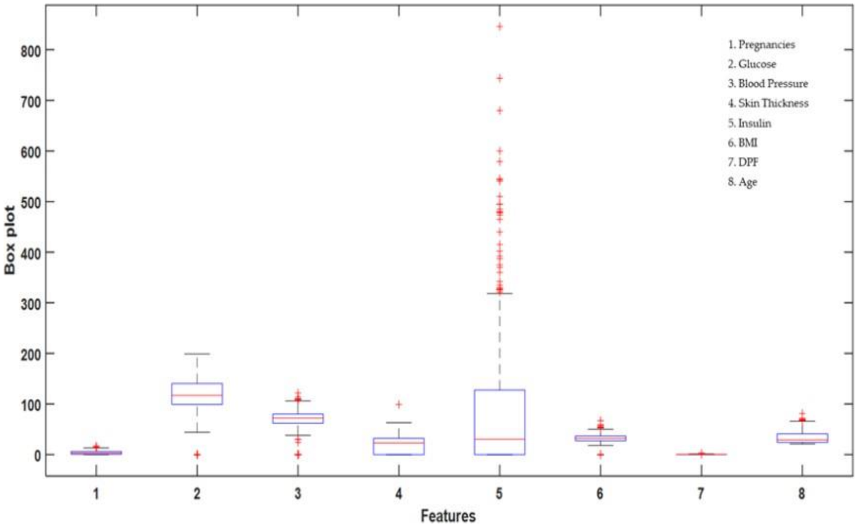


Figure 4. The box plot for the feature.

Result and Discussion

To measure the accuracy of the proposed system, we used the following parameters:(a) true positive (TP), (b) true negative (TN), (c) false negative (FN), and (d) false positive (FP). By using these parameters, we define the other performance measures as follows (Islam, R. and Tarique, M., 2022; Rangayyan, R. M., 2015, Jiaa, Y. and Du, P., 2016, & Islam, R. and Tarique, M., 2024).

Accuracy is the most fundamental performance measure, which is simply the ratio of correctly predicted observations to the total number of observations. The accuracy is defined by,

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \qquad (1)$$

Precision or PPV (positive predictive value) is the ratio of correctly predicted positive observations to the total predicted positive observations. The precision is defined by,

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (2)$$

Specificity is a performance metric that measures a model's ability to correctly identify negative instances, also known as the true negative rate, and is defined as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \qquad (3)$$

Recall measures a classification model's ability to find all relevant positive instances in a dataset and is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4}$$

F1 Score is the weighted average of the precision and recall. Therefore, this score accounts for both FP and FN. The F1 score is defined as

$$\text{F1 Score} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall+Precision}}$$

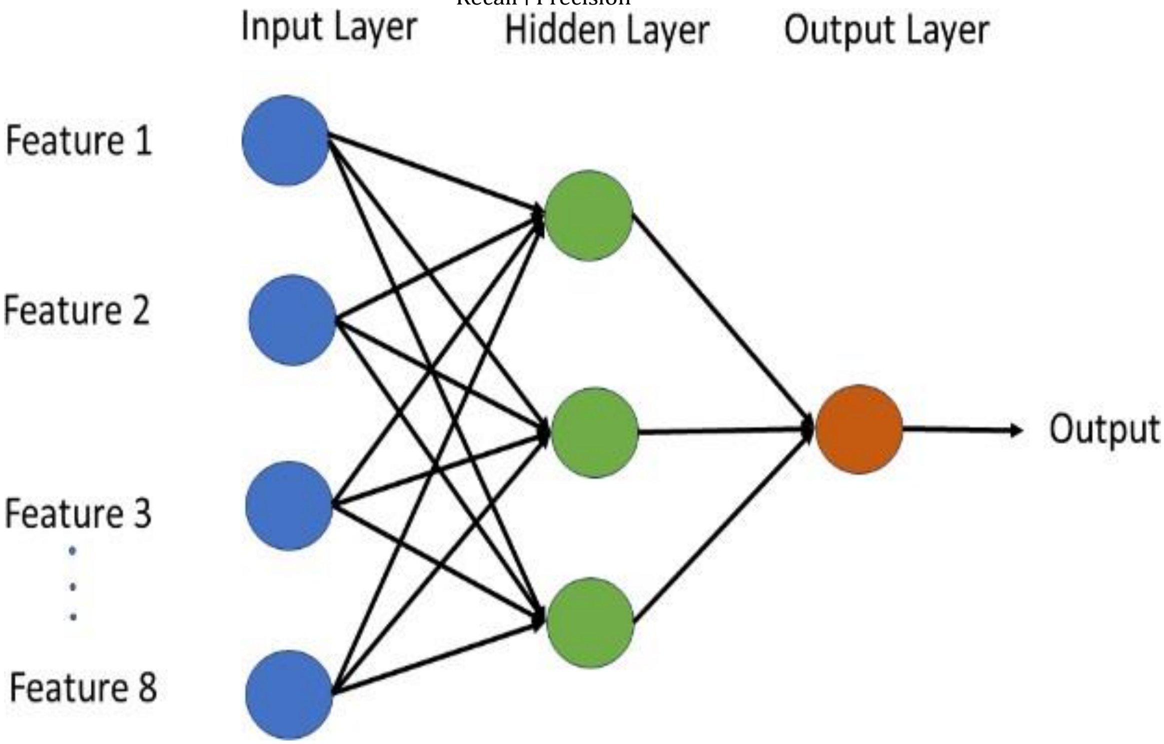


Figure 5. The Feed Forward Neural Network.

In this work, all available samples in the PIMA database have been considered, as mentioned above. Among these samples, 80% of the data were used for training, and the remaining 20% were equally distributed for validation and testing. The error histogram with 20 bins is shown in Figure 6. It visualizes the distribution of the model's prediction errors. This plot also demonstrates that the model's predictions are aligned with the actual target values. The vertical red line on the x-axis shows an error close to zero, indicating accurate predictions. This figure also demonstrates that the proposed model is a well-performing one, as most errors are clustered around zero. The histogram's spread suggests that the model is consistent and that there are no outliers (no significant deviation from the general trend).

Cross-entropies for training, validation, and testing are plotted in Fig. 7. This figure displays the cross-entropy loss between predictions and targets. This plot also shows the superiority of the proposed model. The cross-entropy on the training, validation, and test sets decreases over epochs and reaches 0.475 around epoch 26. The training stopped at epoch 35 after nine (9) iterations of the best validation period. The system's performance is reasonable, as the final cross-entropy is insignificant. The validation and test cross-entropy have a similar pattern. The continuously decreasing cross-entropy indicates that the model has learned effectively and made more accurate predictions with minimal significant overfitting, as illustrated in Figure 7.

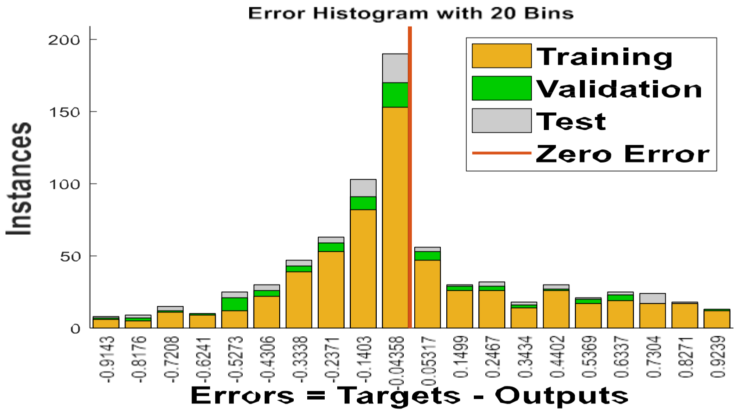


Figure 6. The error histogram plots.

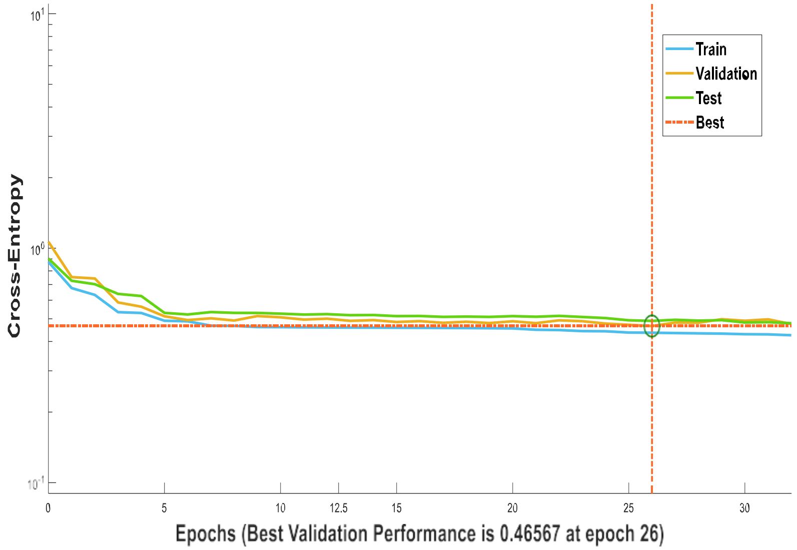


Figure 7. The cross-entropy of training, validation, and testing.

The receiver operating characteristic (ROC) is plotted in Figure 8. It plots the true positive rate (i.e., sensitivity) versus the false positive rate (i.e., specificity) as defined above, with variable threshold values. The ROC curve (bending towards the top-left corner of the graph) indicates a high true-positive rate (TPR) and low false-positive rate (FPR). The perfect test would show points in the upper left corner indicating 100% sensitivity and 100% specificity. The curve is closer to the top-left corner, indicating reasonably good model performance. Table 2 lists the performance measures of the proposed algorithm considering all samples (i.e., training, validation, and testing cases). The proposed model achieved 79.1% accuracy, 62.2% precision, and 73.8% sensitivity.Finally, the confusion matrices for training, validation, and test are shown in Fig. 8. In this figure, “0” indicates patients without diabetes, and “1” indicates patients with diabetes. As mentioned above, 80% of the data samples (i.e., 613) were used for the training. Among these, 433 patients are without diabetes, and 180 patients have diabetes. The confusion matrix for the training is shown in Fig. 8, in the top-left corner. This indicates that the TP and TN were 139 (77.22%) and 351 (81.06%), respectively. These values indicate almost an unbiased training. For validation, 77 samples (52 without diabetes and 25 with diabetes) were used. The confusion matrix indicates that TP and TN were 15 (i.e., 60%) and 44 (i.e., 84.61%), respectively. The validation confusion matrix shows a bias toward classifying patients without diabetes. The confusion matrix of the testing result also indicates some bias. It suggests that the model detects more healthy samples (i.e., 80.70%) than diabetic samples (i.e., 60%). The combined confusion matrix is also shown in the same figure at the bottom-right corner. The green diagonal elements are percentages of correctly classified cases. The corresponding off-diagonal elements are percentages of misclassified cases. It also shows that the model accurately detects 441 healthy samples (81.36%) and 166 diabetic samples (74.77%). Hence, we conclude that the proposed model is a little biased towards detecting healthy patients. The bottom-right cell of all confusion matrices shows the overall correctly predicted classes (in green, %), which is 79.1%, across training, testing, and validation. It also displays the overall misclassified cases, i.e., 20.9% (in red %).

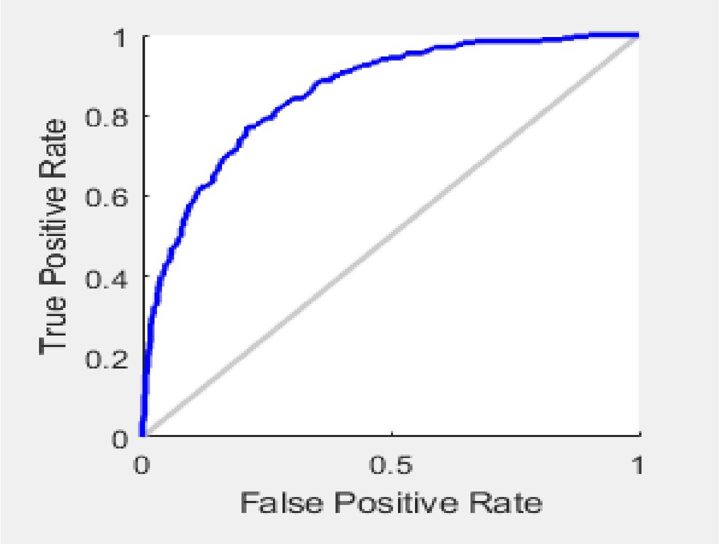


Figure 8. The ROC

Table 2. The Performance Measures

Measures	%
Accuracy	79.1
Precision/ PPV	62.2
Recall/ Sensitivity	73.8
F1 Score	67.5

Specificity	81.4

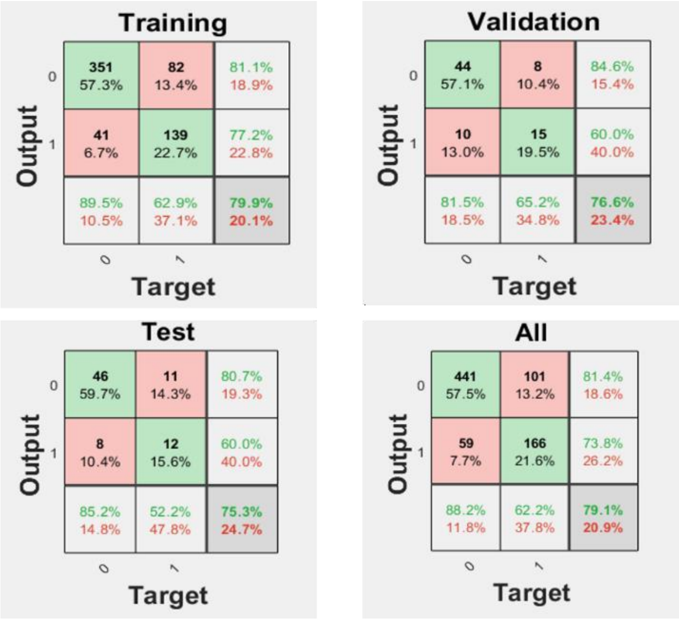


Figure 9. The confusion matrix.

Conclusion

Physiological data contain useful information to identify diabetic patients at an early stage. This investigation shows that tracking physiological data, such as skin thickness, diabetes pedigree function, age, BMI, and the number of pregnancies, alongside monitoring abnormal blood sugar and insulin levels, can help physicians detect diabetes at an early stage. This paper also demonstrates that a shallow FFNN can objectively detect people with diabetes when multiple physiological data are used. The discriminative power of physiological data enabled even a simple and shallow network to perform admirably. Furthermore, such a shallow network significantly reduces the computational burden of the generated system.

This work focuses on binary classification. However, this algorithm can be extended to perform multiclass classification of different types of diabetes, depending on the availability of data. This work only considers the PIMA database. To ensure the proposed algorithm's versatility, other databases, such as the Diabetic Dataset 2019 and the BIT2019 datasets, can also be investigated. This work limits its efforts to implementing the FFNN algorithm. However, other deep learning algorithms, such as CNNs, including transfer learning, should also be investigated to assess the superiority of the proposed algorithm further. Above all, this research anticipates that integrating the knowledge and skills of medical professionals and technologists, along with individuals' awareness, could improve the physical well-being of the general population in the future.

References

Abel, J.J. (1926). Crystalline insulin. Proceedings of the National Academy of Sciences.12(2), 132-136, DOI: https://doi.org/10.1073/pnas.12.2.132

Abousaber, J., Abdallah, H. F., & El-Ghaish, H. (2025). Robust predictive framework for diabetes classification using optimized machine learning on an imbalanced dataset. Frontiers in Artificial Intelligence, 7, article no. 1499530, DOI: https://doi.org/10.3389/frai.2024.1499530

Ahmed, A., Khan, J., Arsalan, M., Shahat, A. A., Alhalmi, A. & Naaz, S. (2024). Machine Learning Algorithm-Based Diabetes Among Female Population using PIMA Dataset. Healthcare, 13(1), article: 37, DOI: https://doi.org/10.3390/healthcare13010037

Ashour, A. F., Fouda, M. M., Fadlullah, Z. M., & Ibrahem, M. I. (2024). Optimized Neural Networks for Diabetes Classification using Pima Indian Diabetes Database. Proceedings of the 3rd International Conference on Computing and Machine Intelligence, Mt Pleasant, April 13-14, 2024, pp.1-7, DOI: https://doi.org/10.1109/ICMI60790.2024.10585703.

Best C. H., & Scott, D. A. (1923). The Preparation of Insulin. Journal of Biological Chemistry. 57(3), 709–723, DOI: https://doi.org/10.1016/S0021-9258(18)85482-5

Chang, V., Bailey, J., Xu, Q. A., & San, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, 35, 16157-16273, DOI: https://doi.org/10.1007/s00521-022-07049-z

Crowfoot, D. (1935). X-ray Single Crystal Photographs of Insulin. Nature. 135, 591–592, DOI: https://doi.org/10.1038/135591a0

de Leiva-Hidalgo A, & de Leiva-Pe´rez A. (2023). On the occasion of the centennial of insulin therapy (1922–2022), II-organotherapy of diabetes mellitus (1906–1923): acomatol. Pancreina. Insulin. Acta Diabetol. 60(2), 163–189, DOI:https://doi.org/10.1038/135591a0

Deepalakshmi, M., Deppalashmi, P., & Nagaraj, P. (2025). Efficient Diabetes Detection using Ensemble Models Optimized by Firefly Algorithm. Proceedings of the second International Conference on Cognitive Robotics and Intelligent Systems, Coimbatore, June 25-26, pp. 573-578, https://doi.org/10.1109/ICC-ROBINS64345.2025.11086333.

Diabetic Dataset 2019 (2025, September 28) Available at https://www.kaggle.com/datasets/tigganeha4/diabetes-dataset-2019

Gunther, A. (2025, August 21) “What is the normal range for blood sugar?” Available at https://www.ynhhs.org/articles/what-is-healthy-blood-sugar

Hounge, P. & Bigirimana, A. G. (2022). Leveraging Pima dataset to Diabetic Prediction: Case Study of Deep Neural Network. Journal of Computer Communication, 10(11), 15-28, DOI: hppts://doi.org/10.4236/jcc.2022.1011002

Huynh, H. V. T., Nguyen, N. H., Qiao, R. (2024). Performance Analysis of Diabetes Detection using Machine Learning Classifier. International Journal of Management and Data Analytics, 4(1), 43-54

International Diabetes Federation(2025, August 15). Diabetes Around the World-2024. available at https://diabetesatlas.org/media/uploads/sites/3/2025/04/IDF_Atlas_11th_Edition_2025_Global-Factsheet.pdf

Islam, R., & Tarique, M. (2022). Chest X-ray Images to Differentiate COVID-19 from Pneumonia with Artificial Intelligence Techniques. International Journal of Biomedical Imaging, 2022(1), article: 5318447, 1-15, DOI: https://doi.org/10.1155/2022/5318447

Islam, R., & Tarique, M. (2024). Artificial Intelligence (AI) and Nuclear Features from the Fine Needle Aspirated (FNA) Tissue Samples to Recognize Breast Cancer. Journal of Imaging, 10(8), Article ID: 201, 1-14,

Islam, R., Abdel-Raheem, E., & Tarique, M. (2022). Voiced Features and Artificial Neural

Network to Diagnose Parkinson's Disease Patients. Proceedings of the International Conference on Electrical and Computing Technologies and Applications, November 23-25, American University of Ras Al Khaimah, United Arab Emirates, 132-136, ID: https://doi.org/10.3390/jimaging10080201

Islam, R. & Tarique, M. (2024). Spectrogram and Mel-Spectrogram Based Dysphonic Voice Detection Using Convolutional Neural Network. Proceedings of the International Conference on Electrical, Computer and Energy Technologies (ICECET, Sydney, Australia, 2024, pp. 1-5, DOI: https://doi.org/10.1109/ICECET61485.2024.10698112.

Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. ICT Express, 7(4),432-439, DOI: https://doi.org/10.1016/j.icte.2021.02.004

Liu, B., Song, L., Zhang, L., Wang, L., Wu, M., Xu, S., Cao, Z., & Wang, Y. (2020). Higher Numbers of Pregnancies Associated With an Increased Prevalence of Gestational Diabetes Mellitus: Results From the Healthy Baby Cohort Study. Journal of Epidemiology, 30(5), 208-212, DOI: https://doi.org/10.2188/jea.JE20180245

Malakar, S., Roy, S.D., Das, S., Sen, S., Velásquez, J.D., & Sarkar, R. (2022). Computer-Based Diagnosis of Some Chronic Diseases: A Medical Journey of the Last Two Decades. Archives of Computational Methods in Engineering. 29(7), 5525-5567, DOI: https://doi.org/10.1007/s11831-022-09776-x

Mousa, A., Mustafa, W., Marqas, R. B., & Mohammed, S. H. M. (2023). A Comparison study of Diabetes Detection using the PIMA Indian Diabetes Database. Journal of the University of Duhok, 26(2) 277-288, DOI: https://doi.org/10.26682/sjuod.2023.26.2.24

Naz, H. & Ahuja, S. (2020). Deep Learning Approach for Diabetic Prediction using PIMA Indian Dataset. Journal of Diabetic and Metabolic Disorders, 19(1), 391-403, https://doi.org/10.1007/s40200-020-00520-5

Petra, R., & Khunba, B. (2020). Analysis and Prediction of Pima Indian Diabetes Dataset using SDKNN Classifier Technique. Proceedings of International Conference on Recent Trends in Engineering and Technology, Tamil Nadu, December 4-5, 2020, DOI 10.1088/1757-899X/1070/1/012059

Pima Indians Diabetes Dataset (2025, August 18) available at https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

R. M. Rangayyan (2015). Pattern Classification and Diagnostic Decision. in Biomedical Signal Analysis. Second Edition, John Wiley and Sons, 111 River Street, NJ, 598-606

Sanger, F. (1960). Chemistry of Insulin: Determination of the structure of insulin opens the way to greater understanding of life processes. Science. 129(3359), 1340-1344, DOI: https://doi.org/10.1126/science.129.3359.1340

Tasnin, I., Nabil, T. U., Islam, S. & Khan, R. (2022). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10(1-2), 1-10, DOI: https://doi.org/10.1049/htl2.12039

Toleva, B., Atanasov, I., Ivanov, I., & Hooper, V. (2025). An Effective Methodology for Diabetes Prediction in the case of Class Imbalance. Bioengineering, 12(1), Article:35, 1-1, DOI: https://doi.org/10.3390/bioengineering12010035.

Vecchio, I., Tornali, C., Bragazzi, N.L., & Martini, M. (2018). The discovery of insulin: an important milestone in the history of medicine. Front Endocrinol (Lausanne). 9(613), 1-8, DOI: https://doi.org/10.3389/fendo.2018.00613

World Health Organization (WHO) (2025, August 16). Diabetic Fact Sheet. Available at https://cdn.who.int/media/docs/default-source/searo/nde/sde-diabetesfs.pdf?sfvrsn=7e6d411c_2

Jiaa, Y., & Du, P. (2016). Performance measures in evaluating machine learning-based bioinformatics predictors for classifications. Quantitative Biology, 4(4), 320-330, DOI: https://doi.org/10.1007/s40484-016-0081-2

Zarger, O. S., Bhagat, A., & Teli, T. A. (2024). A Deep Learning Based Diabetes Diagnosis Model on PIMA Image Dataset. Journal of Electrical Systems, 20(3), 1276-1289, DOI: https://doi.org/10.52783/jes.1444.