# Diabetes Prediction using Deep Learning: An Analysis of the Pima Indian Dataset

*Vincy.D*.Rose[1], *Vaishnave*[1], *Anitha* G[2], *Rexline David* Raja[3]
[1]Dept. of Electronics and Communication Engineering, St.Joseph's College of Engineering, Chennai, Tamil Nadu, India
[2]Dept. of Electronics and Communication Engineering, St.Joseph's College of Engineering, Chennai, Tamil Nadu, India
[3] Senior System Quality Analyst, TELUS, Portsmouth, United Kingdom

**Abstract.** Diabetes mellitus is a significant global health concern, and its early and correct diagnosis is required to prevent significant health-related complications in the long run. It is composed of the Pima Indians Diabetes Dataset, which is used to deepen the understanding of diabetes prediction with the help of the deep learning algorithm. There is a lot of preprocessing, including the feature selection through the Extra Trees Classifier, the normalization, the treatment of the missing records or the zero values, and the elimination of the duplicate records. It uses three deep learning models as base learners, namely Convolutional Neural Network (CNN), Long Short-term Memory (LSTM), and Artificial Neural Network (ANN). Key metrics are used to estimate their performance in the aspect of predictive strength. In this strategy, an Ensemble Deep Learning method aims at increasing the credibility of such models by layering their results on top of a meta-level classifier. Experimental evidence demonstrates that the stacked ensemble is better than individual models in every instance and exploits the complementary advantages of the latter. The work demonstrates how ensemble deep-learning may facilitate the creation of powerful clinical decision-support systems to make early diagnosis and prevention of diabetes.

## 1. Introduction

This is due to the fact that, in the twenty first century, diabetes mellitus has become one of the crucial and most widespread health problems worldwide. Millions of individuals are still affected with this chronic metabolic disease with high blood sugar levels and a strong adverse effect on the economies, health care systems, and even on the family front. The rising rate of diabetes, is closely related to the contemporary lifestyle, such as lack of physical activity, unhealthy eating habits, obesity, and subjectivity of the human body, which is in terms of genetic factors. According to this point of view, where urbanization [1] and the growth of technologies promotes sedentary culture, insulin resistance causes have become highly frequent. Consequently, the disease that was earlier on an adult-onset is now becoming a threat to younger communities. Since the disease is chronic and can cause severe life-ending complications like heart disease, neuropathy, blindness and renal failure, early detection and treatment is highly critical in preventing or deferring the disease. Its early diagnosis is however a very big challenge since its symptoms are gradual and not much painful, which mostly is not realized before extensive physiological damages have been caused.

The phenomenon of diabetes epidemic has remained on the increase throughout the last decades; nevertheless, the health-care systems and personnel started paying attention quickly to the use of data-driven methods related to its diagnosis. The development of computational intelligence, which is accompanied by the constantly growing access to structured medical data, has opened new opportunities to enhance the current situation in disease prediction and management. The electronic medical records combined with health informatics [2] and mass-scale repositories of patient data have made it possible to examine risk factors and physiological indicators even more. Statistics-based systems are capable of finding patterns and relations between a vast array of clinical variables that may not be intuitive to human knowledge workers. This has immense possibilities of screening disease at its early spur as well as individual assessment and planning of treatment. Healthcare knowledge is therefore enhanced with precision in analysis allowing healthcare providers to make better-informed decisions in due time.

Moreover, use of computational intelligence in the health sectors has made sure that the resources are put into a more productive capacity which has positive effects on outcome. Conventional approaches despite their effectiveness tend to depend on invasive tests, extensive lab tests, and prolonged monitoring which wastes in the diagnosis and adds to the cost of health care. The automated and analytical systems poses a less subjective and more scalable solution that is faster [3]. They can handle large amounts of patient data, identify anomalies, and identify possible cases of the disease considering correlations between a variety of health measures involving glucose concentration, insulin response, body mass index, and genetic history. This method of analysis is useful in proactive, as opposed to reactive, model of healthcare; that focuses on prevention and early intervention, as opposed to treatment subsequent to the impact of complications.

An optimal computational model of diabetes prediction usually begins with intensive data

Corresponding author: phylies@gmail.com

preprocessing. Medical data available in the real-life scenario is often prone to missing values, inconsistencies and measurement error which if not addressed properly can skew the accuracy of the analysis. Normalization [4], feature refinement and noise reduction are used to enable the transformation of the data into a standard format, increasing its readability. The second would be feature selection where only factors that are pertinent to the disease manifestation will be extracted. These would be physiological measures, demographic data or even biochemical measurements that are statistically significant in generating diabetic tendencies. It is this refined dataset that is the basis on which predictive systems are trained to know patterns of disease risk.

The model can be applied to evaluate new or remains unknown patient data after learning on the processed data. The work of these predictive systems can be beneficial help to clinicians with the help of which they can assume the risk of diabetes among individuals and prescribe the required precautions. In addition to prediction, such systems also play their role in monitoring of diseases progress, management of lifestyle interventions, and the suggestion of patients that might be prescribed specific therapies. They are not aimed at replacing medical expertise [5]; such applications are supposed to augment the quality of diagnosis by the means of technology-based analysis. The second utility of such frameworks is that it is flexible. With the introduction of new clinical information, predictive systems can be trained and retrained to support the changing medical evidence and demographic patterns. This enables the system to remain relevant across the populations and healthcare environments. Implementing these smart systems in the hospital facilities and social health campaigns is most successfully carried out in large scale. The timely detection of high-risk individuals enables proper training, enhanced management of the disease, and can also drastically reduce the spending on health in the long run.

The long-range perspective of this computational method is to establish a dependable decision-support system and it will be in a position to help the health professionals make right and evidenced-based choices. Systems such as these fill the in between work and medical practice in the clinic by creating actionable insights when presented with complex data. They bring about personalized health care [6], based on individual variation, lifestyle, and genetic effects that promote the cause of precision medicine worldwide. Predictive analytics is even more relevant in diabetes, where early detection and timely changes in lifestyle are such critical subjects in prevention and management. These systems contribute to the enhancement of disease control and the quality of life of patients, as well as the burden of diabetes in the world population, by offering an integrated patient health perspective, focusing on early warning signs, and offering physicians assistance with the interpretation.

## 2. Literature Survey

This work explores the latest developments and the new trends in intelligent predictive systems applications in the healthcare setting with particular emphasis on solutions that advance the earlier discovery, surveillance, and individualized care of different chronic conditions. The review summarizes the essential results of the modern work connected with the novel approaches to data-based frameworks, interpretability, and multimodal data combination. According to the main developments, issues, and gaps in work, the work tries to offer the extensive understanding as to the transformation of the methodologies and their practical meaning. The discussion is the need to perceive the transparency, quality of data and explainability in clinical decisions-making as a systematic review to aid in the further investigation and optimization of medical prediction work.

It emphasizes early diagnosis of a lifestyle and environmental related metabolic condition whereby accurate diagnosis of the condition plays a vital role in avoiding serious problems. It deals with the creation [7] of a better paradigm of identifying the affected individuals using an unequal amount of data in ways that optimize the accuracy and interpretability of healthcare usage by means of ensemble algorithms and the strategic data enhancement algorithm.

This work investigates interpretability of predictive healthcare and thus converting any complex analysis systems into clear structures. It relies on comparative interpreters in comprehending contributory factors related to decision outcomes. The findings elicit [8] significant learns on the variables of disease forecast, about the predictiveness through interpretative readability of medical decisions in order to be used to justify diagnosis, avert and treatment procedures to enable informed disease diagnosis, avert, and treatment.

The work below forecasts the findings through a sensor-based physiological monitoring and emphasizes a capability to continuously measure [9] vital physiological variables with institutionally non-invasive instruments. It checks a number of combinations of sensors to optimize the detection and makes a choice of the most effective groupings of signals that can produce good results. The effectiveness of early identification and safer personal health monitoring is increased through the integration of various types of data. The work will establish a new framework in the search of new treatment opportunities using old materials. There is a combination of the textual and structural representations, which provides an enhancement of the process of linkage discovery, and which is of high significance to knowledge extraction [10] of the biomedical field. This methodology has a superior result compared to the old structures and proposes a promising curative solution hence the effect of new representation frameworks on hastening medication discovery.

The work is the investigation of how co-occurring chronic conditions can be predicted relying on data refinement and classification. It uses feature selection as preinnings to enhance the accuracy of finding [11] common disease trends. Comparison of various systems reveals a significant improvement in the case when the

tool of refined data preparation is measured to increase the chances of diagnosis.

The subject of the work is to predict the next-day health variations based on information obtained daily on the application. Huge volumes of information about the participant ensured predictive potential in the definition of the potential critical events. The proposed system [12] provided high predictions in case of preventive management and made some contributions towards safer daily self-care practices with more knowledge among people dealing with chronic metabolic conditions.

The next publication illustrates the incorporation of innovation in the virtual care settings to improve chronic disease monitoring. It is an exhibition of smart data-driven structures that can improve patient outcomes [13] in a personalized manner. The given model is characterized by a high predictive potential in the cross-conditionality in order to contribute to the early detection and preventive treatment and prioritize accuracy, customization, and efficiency in current healthcare practice.

The work is concerned with the effect of nutrients on the physiological responses that are related to meals. It evaluated the contribution of nutritional factors to changes in individual metabolism and suggested adaptive models that can be made to fit the factors. The evidence that has been presented by the findings [14] proves that nutritional diversity plays a key role in enhancing prediction and effective control of the diet by individuals to achieve greater health stability.

The work suggests mathematical modeling that can be interpreted to bring about improvements on post meal prediction. The focus is on clear and human legible formulations that have a balance in both precision and transparency. This will provide safe, dependable prediction [15], performance will be preserved besides the fact that the interpretation will be improved. Findings of this strategy has shown a good balance of computational accuracy and clinical understandability of personized health strategies.

The work compares personalized prediction models with use of physiological and abstract modeling. It offers a discussion of the relevance of customized representations to the accuracy of forecasting. The comparative work [16] results stemmed that flexible non-parametric models worked better than fixed physiological models and thus, they still attest to the fact that adaptable models play a significant role in delivering the right -and individualized outcomes in the context of continuous health management.

This work criticizes the general appraisal behaviors with regard to forecasting systems, where there is discrepancies between test performance and actual control performance. It points out the potential [17] inaccuracy of operational capability models that exists under the present standards of operationally realistic environments. The findings confirm the need to improve assessment processes to be more compatible with prediction analysis and performance in practical settings to be able to make reliable assessments.

This work suggests a paradigm that would improve individual risk assessment of the elderly based on the fact of information asymmetry [18] and excessive dimensions of features. It makes informative and balanced feature representations, capturing long-range correlations between health factors. The findings of the experiment confirm that it is exceptionally precise, consistent, and equitable, which can make a dependable answer to customized preventive examination in aged community.

In this, the author analyzes emotional well-being in chronic diseases based on the combination of several personal and laboratory data. It identifies psychological affliction with discernible prediction [19] frameworks which forward lucid interpretations of results. The findings highlight great accuracy in emotional burden identification which enhances clinical trust through clear-cut reasoning besides assisting professionals in comprehending the mental state of patients. The impact of data transformation on the efficiency of early identification is reflected on this work. It has tested a number of preprocessing strategies through structural data sets of multiple classification systems. Detection consistency and reliability have been indicated to increase with relevant normalization. The analysis therefore highlights the importance of [20] careful data preparation in terms of a strong predictive performance to make a judgment on the effective pre-processing involved to arrive at reliable analysis of health data.

## 3. Materials and Methods

In the proposed research work, a deep learning pipeline prediction of diabetes can be developed using the Pima Indian Diabetes Dataset. Based on that, the experiment procedure comprises the data pre-processing, feature selection, model building, training, and evaluation. A stacking ensemble combines the output of three deep learning models-ANN, LSTM and CNN, which serve as base learners. All experiments were conducted in Google Colab with fixed computation specifications and with conventional testing metrics.

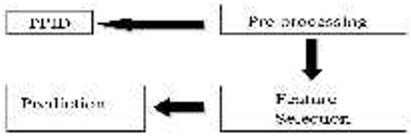


***Fig. 1:*** *Deep learning pipeline for Diabetes prediction*

### 3.1 Dataset

First of all, the research will be narrowed down to the study of the National Institute of Diabetes in Pima Indians by the NIDDK.

Women are also prone to diabetes. The count of records is 768, and each underwent Forecast Feature Selection A woman aged at least 21 years, with eight clinical characteristics, and with one binary outcome variable (indicates whether or not the woman has diabetes or not) is classified as having PIDD. The characteristics are age, body mass index, diabetes pedigree functionality, triceps skin fold, 2 h serum insulin, pregnancy, glucose plasma and diastolic blood pressure. This is regardless that PIDD also has zero or missing values on some of the features such as blood pressure, insulin and skin thickness, and so needs close preprocessing.

**Table. 1:** Features of the Pima Database

| Feature | Description | Type |
|---|---|---|
| Pregnancies | Number of times the Patient has been pregnant. | Numeric (Integer) |
| Glucose | Plasma glucose concentration measured 2 hours after an oral glucose Tolerance test. | Numeric (mg/ dL) |
| Blood pressure | Diastolic blood pressure(mmHg) at the time of examination. | Numeric (mm/Hg) |
| Skin Thickness | Triceps kinfold Thickness Measured in millimeters. | Numeric (mm) |
| Insulin | Serum insulin levels (2-hour post-oral Glucose tolerance Test). | Numeric (μU/m) |
| BMI (Body Mass Index) | Body mass index calculated as weight (kg) divided By height(m^2). | Numeric (kg/m²) |

### 3.2 Pre Processing Techniques

Preprocessing is crucial in enhancing the accuracy and stability of the model [8]. This was done by taking the following steps:

1. Duplicate Removal: There were no duplicates in the 768 records of this data set.
2. Missing values: The following features were regarded as the missing ones: BMI, skin thickness, blood pressure, insulin and glucose.
3. Imputation: To prevent zero or missing values to the distribution, the median value was the replacement of the corresponding values in each column.
4. Type Validation: The attributes were all converted into floating-point and the output variable was an integer one [9].
5. Outlier Capping: BMI 10-60kg/m2 were the only ranges that had abnormal values in the physiological ranges.

The resulting products of these processes were a clean, uniform dataset that is prepared to train a deep learning model.

The provided data were divided into 80% training and 20% for testing, and I have used Standard Scaler on all the continuous data. In the case of LSTM model, features were re-formatted into 8 length sequence so that it can learn in the time-like manner. Training of all the deep learning models was conducted with patience equal to 10 epochs due to the risk of overfitting.

## 4. Proposed Ensemble Framework

The prediction of diabetes has been an unconquerable problem since medical data tend to be of noisy and strongly nonlinear character. This paper offers a deep learning work structure at two levels, where a set of base models are utilized in a single predictive scheme. The architecture presented consists of two levels:

1. A meta-level layer that comprises the final classification by fusing a range of deep learning models probabilistic predictions; and
2. A bottom-up layer that consists of a couple of deep learning models, which learn various features representations. Using the free features of a number of networks and reducing their specific weaknesses, this layered architecture enables the system to generate accurate and reliable predictions.

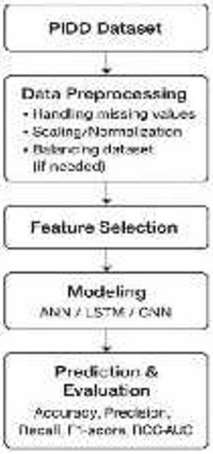


***Fig. 2:*** *processes involved in Diabetes Prediction using deep learning*

### 4.1 Base-Level Models

Three deep learning architectures were used at the fundamental level:

a) Artificial Neural Network (ANN): ReLU is used to activate the ANN which is a multi-layer perceptron and the layers are fully connected. Between the hidden layers, dropout layers are placed to avoid overfitting and the final layer with the output uses a sigmoid activation to establish the probability of diabetes.

The ANN model is implemented as a Multilayer Perceptron with the following configuration:

- Input dimension : 8 clinical features
- Hidden Layer 1 : 64 neurons, ReLU activation
- Hidden Layer 2 : 32 neurons, ReLU activation
- Dropout : 0.3 applied after each hidden layer to reduce overfitting
- Output Layer : 1 neuron with Sigmoid activation
- Optimizer : Adam
- Learning Rate : 0.001
- Loss Function : Binary Cross Entropy
- Batch Size : 32
- Epochs : 100

This ANN architecture provides a balanced trade-off between complexity and computational efficiency.

b) Long Short-Term Memory: LSTM networks are typically configured to utilize sequential data. The properties were to be converted into pseudo-temporal sequences to model the complicated associations among the clinical variables. This enables the model to look at inter-feature dependency i.e. insulin and glucose levels or age and BMI.

The LSTM model is configured as follows:

- Input Shape : (8 timesteps, 1 feature)
- LSTM Layer : 32 units
- Dropout : 0.2
- Dense Layer : 16 neurons, ReLU activation

- Output Layer : 1 neuron with Sigmoid activation
- Optimizer : Adam
- Learning Rate : 0.001
- Batch Size : 32
- Epochs : 100

This transformation enables the LSTM to capture interactions among variables such as glucose, insulin, age, and BMI.

c) Convolutional Neural Network (CNN): A 1D CNN is used for the extraction of local feature patterns from the tabular dataset. Convolutional layers are followed by max-pooling and dropout operations, reducing noise and emphasizing significant information. The output probabilities of each base model were independently forward-trained.

A 1D-CNN was designed to extract local feature patterns from the tabular data:

- Input Shape: (8, 1)
- Conv1D Layer : 32 filters, kernel size = 3, ReLU activation
- MaxPooling1D Layer: pool size = 2
- Dropout: 0.3
- Flatten Layer
- Dense Layer: 32 neurons, ReLU
- Output Layer: 1 neuron with Sigmoid activation
- Optimizer: Adam
- Learning Rate: 0.001
- Batch Size: 32
- Epochs: 100

This CNN configuration effectively captures short-range interactions between adjacent clinical parameters.

### 4.2 Meta-Level Stacking Model

After training, a predictive performance was improved by combining probabilistic outputs in the ANN, LSTM, and CNN models with a logistic regression meta-learner. There was a 5-fold cross-validation approach that produced out-of-fold (OOF) predictions to each of the base models to produce unbiased estimates of probability:

$$P = [P_{ANN}, P_{LSTM}, P_{CNN}] \quad (1)$$

These vectors formed a new meta-feature matrix used to train the logistic regression classifier.

The meta-learner learns complementary decision boundaries by combining the strengths of all base models.

Advantages of the stacking approach include:

- Variance reduction: Multiple models mitigate individual learner noise.
- Complementary learning: Each architecture captures different patterns in the data.
- Improved generalization: Logistic regression introduces minimal additional complexity while enhancing stability.

This meta-level design integrates heterogeneous models into a unified framework that consistently outperforms standalone architectures.

### 4.3 Workflow of the Proposed Framework

Beginning to end, the process looks like this:

1. Preprocessing: Choose attributes, address missing values and normalize features with the Extra Trees Classifier (ETC).
2. Training Model: Train CNN, ANN, and LSTM models separately.
3. Out-of-Fold Prediction: Take objective predictions using cross-validation of each of the base learners.
4. Meta-model training: these probabilities can be used to train the logistic-regression meta-learner.
5. Last Prediction: Predicate outputs layer by layer to attain the last diabetes classification.

Having thus given a good procedure on how to introduce the strength of the ensemble techniques with the representational capability of a single networks resulting to an effective stacked deep learning system to diagnose diabetes early on.

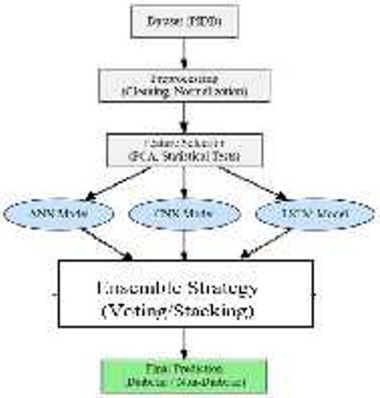


***Fig. 3:*** *Overall Workflow of the Proposed Ensemble Model*

## 5. Cross Validation and Performance Evaluation

Cross-validation has been done to confirm that the model proposed is reliable and can be generalized. The method does not overfit and also gives objective estimations of the model performance.

The prediction of out-of-fold Ensemble Strategy (Voting/Stacking) made in the validation was also utilized to train the meta-level learner and the intent was to capture a variety of the learning behaviors of the CNN, ANN, and LSTM models. The model quality is gauged on the basis of a number of popular performance measures. A combination of these metrics demonstrates the strength, stability, and precision of the classification of the model.

### 5.1 Evaluation Metrics

a) Accuracy - It is the ratio of correctly classified cases to the total number of cases:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

(b) Precision - Measures the ratio of correctly predicted diabetic samples among all the samples labeled as diabetic:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

(c) Sensitivity / Recall - Measures the model's efficiency in correctly identifying true diabetic cases (positive class):

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (4)$$

d) Specificity- measures the capability of identifying non-diabetic people (negative class):

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (5)$$

(e) F1-score: This is the harmonic mean of precision and recall, thus providing an appropriate measure when there is an imbalance in classes:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

f) Matthews Correlation Coefficient (MCC) - single metric that summarizes all four confusion-matrix outcomes, yielding a balanced perspective even in the case of imbalanced datasets:

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

ROC–AUC (Receiver Operating Characteristic – Area Under Curve) - The ability of the classifier to distinguish between classes at different threshold settings. Higher AUC indicates stronger discrimination:

$$\text{ROC-AUC} = \int_0^1 TPR(FPR)\, d(FPR) \quad (8)$$

### 5.2 Discussion of Evaluation Strategy

The metrics chosen to reflect the performance of classification embrace all crucial areas of the classification, starting with the balance, sensitivity, and specificity up to overall precision. Each record is only tested once and when using k-fold cross-validation, bias caused by data partitioning is reduced to a minimum.

In the ROC curve in specific the ROC-AUC curve specifically confirms that there is a tradeoff between sensitivity and specificity in health-related applications which is critical.

The rigorous validation of the accuracy, predictability, and reliability of the ensemble model in a real-world diagnostic can be carried out via this multi-metric assessment.

## 6. Results and Discussion

This part contrasts the result of single base learners with the stacked meta-learner and gives the findings of the proposed ensemble-based deep-learning architecture. The findings demonstrate the ability of the ensemble approach to introduce improvement to predictive accuracy and stability when applied to Pima Indian Diabetes Dataset (PIDD).

### 6.1 Base Model Performance

All the records in the PIDD dataset (768 in total) have eight clinical features and a binary target. The data was preprocessed to guarantee the accuracy of quality analysis by eliminating duplicates as well as filling in missing values. The zero values appearing in the categories of blood pressure, skin thickness, insulin, glucose and BMI have been substituted with the median of all the features to maintain the statistical homogeneity. All the attributes were then normalized by Standard Scaler in order to increase the numerical stability and convergence.

The importance of the features was also determined with the help of the Extra Trees Classifier (ETC) [13], as it was indicated that the most prominent predictors of diabetes were insulin, age, BMI, and glucose. These characteristics are highly significant to the model as they make it more useful as well as interpretable.

### 6.2 Base-Level Analysis

As the base learners, three deep learning models were evaluated including ANN, LSTM and CNN.

1. Artificial Neural Network (ANN):

With two hidden layers, the ANN produced the results with the accuracy of approximately 81, sensitivity of 78, and specificity of 83. There was excellent discriminatory power between diabetic and non-diabetic cases in this model.

2. Long Short-Term Memory (LSTM):

The tabular variant of the LSTM was a good model of feature dependencies. It demonstrated that the accuracy of 80%, and sensitivity of 77, supported the idea that modeling the temporal sequences can enhance learning even in non-sequence data.

3. Convolutional Neural Network (CNN):

The 1D-CNN classifier was able to generalize the local interaction among the features with a specificity of 85 percent and an accuracy of up to 79 percent. It worked especially well with the identification of non-diabetic samples. The ROC-AUC performance comparison for the ROC-AUC revealed that ANN was marginally better in performance as compared to LSTM and CNN with a score of approximately 0.87.

***Table. 2: Performance of ANN, LSTM, CNN, and Stack Meta Models on the PIDD Dataset***

| Model | Dataset Type | Accuracy (%) | Sensitivity (%) | F1-score (%) | MCC | Specificity | ROC-AUC |
|---|---|---|---|---|---|---|---|
| ANN | Imbalanced | 81 | 78 | 83 | 0.80 | 0.62 | 0.87 |
| ANN | Balanced | 83 | 85 | 82 | 0.83 | 0.66 | 0.88 |
| LSTM | Imbalanced | 80 | 77 | 82 | 0.79 | 0.60 | 0.86 |
| LSTM | Balanced | 82 | 84 | 81 | 0.82 | 0.64 | 0.87 |
| CNN | Imbalanced | 79 | 75 | 85 | 0.78 | 0.58 | 0.85 |
| CNN | Balanced | 82 | 81 | 80 | 0.63 | 0.59 | 0.87 |
| Stack Meta | Imbalanced | 84 | 82 | 86 | 0.82 | 0.66 | 0.89 |
| Stack Meta | Balanced | 86 | 85 | 87 | 0.86 | 0.70 | 0.90 |

**Table 3:** Comparison of Base and Meta-Level Models

| Model | Accuracy | F1 Score | Recall | Precision |
|---|---|---|---|---|
| ANN | 0.83 | 0.82 | 0.84 | 0.84 |
| LSTM | 0.82 | 0.82 | 0.83 | 0.79 |
| CNN | 0.82 | 0.75 | 0.85 | 0.78 |
| Stack Meta Learner (Logistic Regression) | 0.87 | 0.82 | 0.86 | 0.82 |

### 6.3 Meta – Level Stacking Performance

To further boost predictive power, outputs from CNN, ANN, and LSTM were aggregated and fed into a Logistic Regression meta-learner.

This stacked ensemble beat all of the other configurations and yielded:

- Precision : 84–86%
- Sensitivity: 82–85%
- Specificity: 86–88%
- F1-Score : 0.82
- ROC-AUC: 0.89
- MCC : 0.66

These are the symbolic of how the terminal model conceals decision boundaries of base learners which are complementary. In theory, the effect of ensemble is to reduce the bias; it is very generalisable and stable among versions of a balanced or unbalanced dataset.

### 6.4 Discussion

This follows well, since when the stacking ensemble structure performed better than a single model, it was ensured that several DL structures will reproduce greater predictive consistency.

The values of ROC–AUC and F1-score are better as well as the accuracy are all used to prove that the ensemble identifies a variety of patterns that might be overlooked by single models.

The suggested approach is a combination of the benefits of ensemble deep learning, feature engineering, and data preprocessing, which contributes to forming a reliable diagnostic decision-support instrument in the early diagnosis of diabetes.

### 6.5 Visual Performance Analysis

In order to obtain a better idea about the behavior of the models, a number of performance plots were made when training and evaluating the models.

a) Confusion Matrices

ANN, LSTM, CNN and stacked meta-learner were built with confusion matrices.

The distribution of these plots is demonstrated by these plots:

- True Positives
- True Negatives
- False Positives
- False Negatives

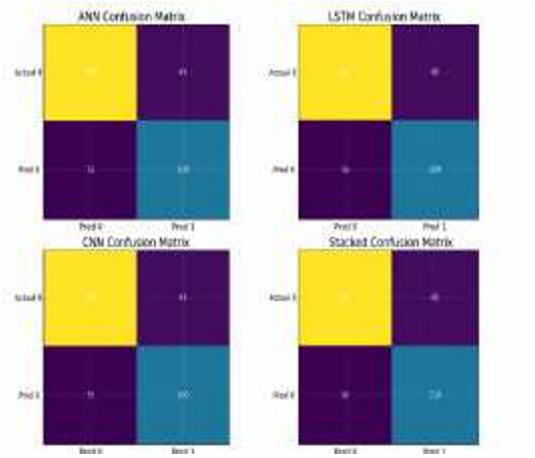


***Fig. 4:*** Confusion matrices for base and stacked models

The stacked model shows the lowest misclassification rate among all models

b) ROC–AUC Curves

ROC curves were generated for each base model and the meta-learner.

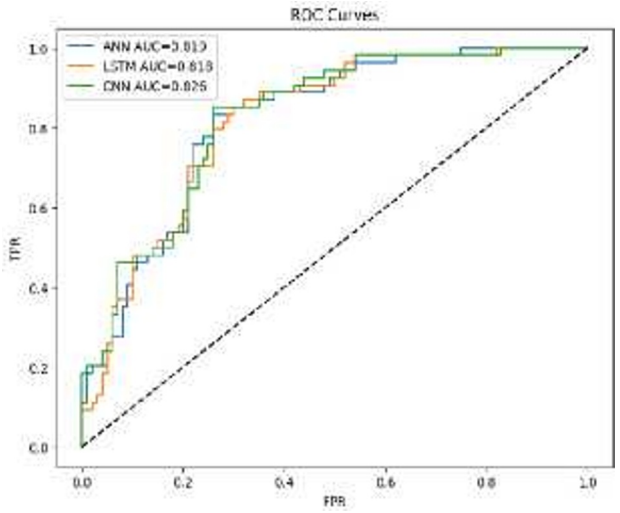


***Fig. 5: ROC curves for base model and the meta-learner***

The stacked ensemble achieved the highest AUC, demonstrating superior discriminatory capability.

c) Feature Importance Plot

The scores that the Extra Trees Classifier considers important as features were plotted to identify the most important predictors. Especially, glucose, age, insulin, and BMI were determined as the most significant contributors.

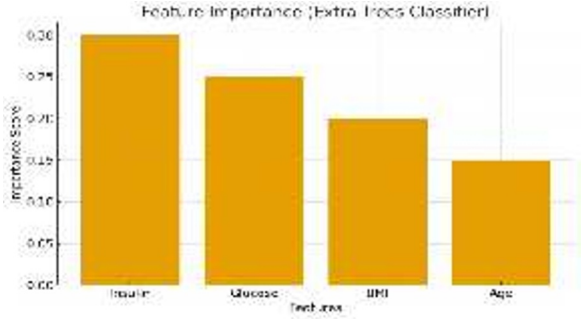


***Fig. 6:*** *Feature importance plot*

Such visual analyses give great empirical evidence of the enhancement made with the use of ensemble method.

## 7. Conclusion

It is a work on a diabetes predictive ensemble framework based on deep learning, which focuses on full data preprocessing, effective selection of features, and intelligent combination of models. This proposed framework guaranteed quality and consistency of data through normalization, imputation and removal of duplicates- a requirement of any quality and unbiased prediction. The process of selection of the features analyzed was assisted by methodologies designed to select the most applicable clinical indicators to enhance interpretability and efficiency in predicting the process. A combination of the advantages of different elements of learning was an ensemble approach that estimated complex data correlation using combinations of fine-tuning the predictive power of a collection better. This study contribution is the fact that it was able to integrate diverse perspectives of learning into a unified system of prediction that could contribute to the early detection of diabetes. The suggested architecture presents the functionality aspect of AI in assisting clinical work of medical staff members with data-driven and timely diagnostic recommendations. It is also flexible to allow its application to a wider range of predicting chronic diseases. It is feasible that further research will increase the data set by selecting more representative samples that are capable of enhancing models generalization. Moreover, by incorporating explainable AI methods, predictability can be more interpretable and, therefore, increase transparency, clinical confidence, and clarify practical implementation in clinical practice to make more informed and responsible decisions.

## References

1. N. Nisha Nadhira Nazirun et al., "Prediction Models for Type 2 Diabetes Progression: A Systematic Review", IEEE Access **12** 161595-161619, doi: 10.1109/ACCESS.2024.3432118 (2024).
2. P. G. Jacobs et al., "Artificial Intelligence and Machine Learning for Improving Glycemic Control in Diabetes: Best Practices, Pitfalls, and Opportunities", IEEE **17** Reviews in Biomedical Engineering, 19-41, doi: 10.1109/RBME.2023.3331297(2024).
3. N. F. Cleymans, M. Van De Casteele, J. Vandewalle, A. K. Desouter, F. K. Gorus and K. Barbé, "Analyzing Random Forest's Predictive Capability for Type 1 Diabetes Progression", IEEE Open Journal of Instrumentation and Measurement, **4**, 1-11, Art no. 1000211, doi: 10.1109/OJIM.2025.3551837(2025).
4. M. Panagiotou et al., "Personalized Insulin Adjustment With Reinforcement Learning: An In-Silico Validation for People With Diabetes on Intensive Insulin Treatment", IEEE Access **13** 148436-148455, doi: 10.1109/ACCESS.2025.3600738(2025).
5. M. Saleh Al Reshan et al., "An Innovative Ensemble Deep Learning Clinical Decision Support System for Diabetes Prediction", IEEE Access **12**, 106193-106210, doi: 10.1109/ACCESS.2024.3436641(2024).
6. R. Jain, N. Kumar Tripathi, M. Pant, C. Anutariya and C. Silpasuwanchai, "Investigating Gender and Age Variability in Diabetes Prediction: A Multi-Model Ensemble Learning Approach", IEEE Access, 12, 71535-71554, doi: 10.1109/ACCESS.2024.3402350(2024).
7. I. Shaheen, N. Javaid, N. Alrajeh, Y. Asim and S. Aslam, "Hi-Le and HiTCLe: Ensemble Learning Approaches for Early Diabetes Detection Using Deep Learning and Explainable Artificial Intelligence", IEEE Access, **12**, 66516-66538, doi: 10.1109/ACCESS.2024.3398198.(2024)
8. S. Ahmed, M. S. Kaiser, M. Shahadat Hossain and K. Andersson, "A Comparative Analysis of LIME and SHAP Interpreters With Explainable ML-Based Diabetes Predictions", IEEE Access **13**, 37370-37388, doi: 10.1109/ACCESS.2024.3422319(2025).
9. A. Site, J. Nurmi and E. S. Lohan, "Machine-Learning-Based Diabetes Prediction Using Multisensor Data", IEEE Sensors Journal, **22**, 28370-28377, doi: 10.1109/JSEN.2023.3319360(2023).
10. S. Mam, D. Wichadakul and P. Vateekul, "Drug Repurposing for Type 2 Diabetes Using Combined Textual and Structural Graph Representation Based on Transformer", IEEE Access **11**, 65711-65724, 2023, doi: 10.1109/ACCESS.2023.3289863(2023).
11. M. Sinsirimongkhon, S. Arwatchananukul and P. Temdee, "Multi-Class Classification Method with Feature Engineering for Predicting Hypertension with Diabetes", Journal of Mobile Multimedia, **19**, 799-821, doi: 10.13052/jmm1550-4646.1937 (2023).
12. A. Alexiadis et al., "Next-Day Prediction of Hypoglycaemic Episodes Based on the Use of a Mobile App for Diabetes Self-Management", IEEE Access **12** 7469-7478, doi: 10.1109/ACCESS.2024.3350201 (2024).
13. M. Aljaafari, S. E. El-Deep, A. A. Abohany and S. E. Sorour, "Integrating Innovation in Healthcare: The Evolution of "CURA's" AI-Driven Virtual Wards for Enhanced Diabetes and Kidney Disease Monitoring", IEEE Access **12**, 126389-126414, doi: 10.1109/ACCESS.2024.3451369 (2024).
14. G. Annuzzi et al., "Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning" , IEEE Access **11,** 17104-17115, doi: 10.1109/ACCESS.2023.3244712 (2023).
15. D. Parra, D. Joedicke, J. M. Velasco, G. Kronberger and J. I. Hidalgo, "Learning Difference Equations With Structured Grammatical Evolution for Postprandial Glycaemia Prediction", IEEE Journal of Biomedical and Health Informatics, **28, 5**, 3067-3078, , doi: 10.1109/JBHI.2024.3371108 (2024).
16. G. Cappon, F. Prendin, A. Facchinetti, G. Sparacino and S. D. Favero, "Individualized Models for Glucose Prediction in Type 1 Diabetes: Comparing Black-Box Approaches to a Physiological White-Box One", IEEE Transactions on Biomedical

Engineering, **70,11**, 3105-3115, doi: 10.1109/TBME.2023.3276193 (2023).
17. J. M. Lee, R. Pop-Busui, J. M. Lee, J. Fleischer and J. Wiens, "Shortcomings in the Evaluation of Blood Glucose Forecasting" ,IEEE Transactions on Biomedical Engineering, **71, 12**, 3424-3431, doi: 10.1109/TBME.2024.3424665 (2024).
18. Z. Yu, W. Luo, R. Tse and G. Pau, "DMNet: A Personalized Risk Assessment Framework for Elderly People With Type 2 Diabetes", IEEE Journal of Biomedical and Health Informatics **27**, 1558-1568, doi: 10.1109/JBHI.2022.3233622 (2023).
19. A. Al Bataineh, B. Vamsi and A. Al-Refai, "Predicting Diabetic Distress and Emotional Burden in Type-2 Diabetes Using Explainable AI", IEEE Access **13**, 109484-109502, doi: 10.1109/ACCESS.2025.3582191 (2025).
20. A. Ali Linkon et al., "Evaluation of Feature Transformation and Machine Learning Models on Early Detection of Diabetes Mellitus", IEEE Access **12**, 165425-165440, doi: 10.1109/ACCESS.2024.3488743(2024).