NPRC Journal of Multidisciplinary Research
Volume 3, Issue 2, February, 2026 Pages: 116-129
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i2.91272

# Ensemble Based Machine Learning Model for Prediction of Diabetes

**Ramesh Prasad Bhatta**
Assistant Professor
Central Department of CSIT,
Far Western University, Mahendranagar, Nepal
rpb.mcs@gmail.com
https://orcid.org/0009-0005-0554-9072

Received: December 29, 2025 Revised & Accepted: February 23, 2026

## Abstract

**Background:** Early prediction of the disease is of prime importance to reduce the risk of complications and health-care expenses, especially in developing countries. Though conventional machine learning algorithms are employed for the prediction of the disease, ensemble learning algorithms are more robust for the same purpose.

**Methods:** The performance of ensemble learning algorithm models such as AdaBoost, Gradient Boosting, XGBoost, and Stacking Ensemble is evaluated in this paper by using the PIMA Indian Diabetes dataset. In data preprocessing, missing value handling, normalization, and splitting are performed. The performance of these models is also evaluated by calculating their accuracy, precision, recall, F1-score, and ROC-AUC scores.

**Results:** Among all the models that were tested, it was observed that the Stacking Ensemble model provided the best results in terms of accuracy, precision, recall, F1-score, and ROC-AUC value, with 0.86 accuracy, 0.82 precision, 0.80 recall, 0.81 F1-score, and 0.91 ROC-AUC value. The XGBoost model was also observed to perform well in terms of accuracy and ROC-AUC value, achieving 0.82 accuracy and 0.88 ROC-AUC value. From the results obtained in the above experiment, it is clear that there is an improvement in the performance of the stacked ensemble method.

**Conclusion:** Ensemble learning helps to improve the prediction of diabetes at early stages. The Stacking Ensemble model achieved the highest performance with 86% accuracy, 0.82 precision, 0.80 recall, 0.81 F1-score, and 0.91 ROC-AUC. This study provides a foundation for future research and development of robust predictive models in diabetes care and prevent from diabetic complications.

**Implications:** The similarity in the results for Precision (0.82) and Recall (0.80) for the Stacking model shows good balance, which is critical since the data set could be imbalanced. Even though the XGBoost algorithm provides good trade-offs in terms of its complexity and accuracy, the results show that for the best results, the complexity of the Stacking Ensemble is required.

**Keywords:** Diabetes, Accuracy, Prediction, XGBoost, Machine Learning, Ensemble Learning

## Introduction

Diabetes mellitus is an increasing global health problem that is characterized by hyperglycemia persisting in time because of defects in insulin secretion, insulin action, or both. The increasing prevalence of Type 2 diabetes mellitus (T2DM) has contributed to a lot of pressure on the healthcare systems across the world, particularly in developing countries. As per the International Diabetes Federation (2021), more than 530 million adults worldwide are currently living with diabetes, and this number is expected to rise dramatically in the coming years. The South Asian region is one of the hotspots of this pandemic, and increasing urbanization, sedentary lifestyle, and diet are some of the factors that have contributed to the early onset of this disease. Countries like Nepal, India, Bangladesh, and Pakistan are witnessing an increasing trend of undiagnosed and prediabetes, who are identified and diagnosed only when complications arise.

The laboratory parameters like glycated hemoglobin (HbA1c), oral glucose tolerance tests (OGTT), and fasting plasma glucose (FPG) are the basis of traditional methods of diagnosis. Although traditional methods are useful, they might not be able to detect individuals who are at risk of developing this disease in the early stages of abnormal metabolism. Moreover, traditional methods are reactive rather than preventive. Machine learning (ML) algorithms have been employed more and more in recent years in prediction problems in the medical field, and this is due to the fact that ML can offer data-driven insights. As ML can combine multiple predictions from multiple base learners, it can be particularly useful in this problem, and thus, ensemble learning algorithms can be employed.

In the context of ensemble learning, the primary goal is to use various models to reduce variance and bias, which in turn enables the model to generalize better. In the context of diabetes prediction, various researchers have used bagging, boosting, and stacking methods. Researchers also utilized multiple factors from large population databases to establish an ensemble model in 2020. They found that this method outperformed single models in separating the different outcome measures. The authors advanced this idea and presented a weighted ensemble of gradient boosting and decision trees for bringing additional knowledge to the South Asian community (Dutta et al.,2022).

In the area of diabetes early detection, literature also reports that the boost methods have potential. The work has contrasted XGBoost, LightGBM and CatBoost, with the demonstration that boosting is the clear winner in terms of classification over classical machine learning methods (Ganie et al., 2023).

Oliullah et al. 's (2024) paper has also demonstrated the possibility of using the ensemble techniques, which can be used with a variety of algorithms and indicate that these methods are superior in achieving high accuracy and low variance. For example, in Nepal, the infrastructure for health care is in the developing phase, with less availability of electronic health records, which may create difficulty in developing large datasets for the training and validation of the predictive models. Additionally, the applicability of computationally expensive ensemble methods in resource-constrained health care needs to be addressed.

There is an urgent need for the development of effective, interpretable, and context-aware ensemble learning models for the early detection of diabetes. Features related to the demographic, lifestyle, and clinical variability of the population in Nepal and the surrounding regions can also be used, which may help in the development of more accurate models. Additionally, the combination of XAI with the ensemble methods may also help in the development of more effective models, as well as in the development of policies in support of the models. The proposed research is based on the development of optimized ensemble models with the help of rigorous evaluation methods for the development of scalable, data-driven decision support systems in emerging countries.

## Literature Review

Ensemble machine learning is now considered a hot topic in early prediction of diabetes. In 2020, Yang T. and colleagues proposed an ensemble method based on non-invasive predictors for early screening of Type 2 diabetes. The authors proposed a simple ensemble method combining Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Random Forest (RF). The proposed method was implemented using NHANES data and yielded an AUC of 0.849, which was superior to individual performance. The proposed ensemble method is thus considered efficient in increasing the robustness and reliability of early prediction of diabetes (Yang et al., 2020).

The study proposed a soft voting ensemble method combining Random Forest, Logistic Regression, and Naïve Bayes. The proposed ensemble method was implemented using the PIMA Indian Diabetes dataset and yielded 79.04% accuracy, which was superior to individual performance. The proposed ensemble method is thus considered efficient in increasing the robustness and reliability of early prediction of diabetes (Kumari et al., 2021).

Dutta A. et al. (2022) developed a weighted ensemble model that combines Decision Tree, Random Forest, XGBoost, and LightGBM for early diabetes prediction in South Asian communities. The proposed ensemble model, which utilized a systematic preprocessing workflow, attained an AUC of 0.832 and accuracy of 73.5%. The authors concluded that the combination of boosting and bagging methods improves the predictive accuracy.

A hybrid super ensemble learning approach that combined Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting with SVM meta-learning. The proposed approach was tested on various datasets, including PIMA and hospital data, and showed a maximum accuracy of 99.6%. The research work proved the efficiency of stacking-based super learners in improving generalization performance (Doğru A. et al. 2023).

Ganie S. M. et al. (2023) presented a review of various boosting methods such as XGBoost, CatBoost, LightGBM, AdaBoost, and Gradient Boosting for diabetes prediction. The best accuracy of 92.85% was obtained using Gradient Boosting on the UCI PIMA dataset. The research work concluded that boosting ensembles are efficient in handling nonlinear relationships in medical datasets.

An ensemble approach that combined Decision Tree, Gradient Boosting, and Support Vector Machine models. The researcher tested its method with the PIMA dataset, which resulted in an accuracy rate of up to 93.2%. It also pointed out the importance of early diagnostic tools in the field of health, especially in the context of the COVID-19 pandemic (Thakur D. et al.,2023).

The stacked ensemble, incorporating Random Forest, XGBoost, NGBoost, Bagging, LightGBM, and AdaBoost algorithms. This stacked model had an accuracy rate of 92.91%, which is higher than the individual algorithms used in the model. This is because the algorithms used in the model are diverse, which is an important factor in enhancing the predictions made by the model. This is an efficient model in the context of early prediction of diabetes, as pointed out by the researcher (Oliullah et al., 2024).

Kawarkhe M. and Kaur P. (2024) have proposed an ensemble classifier that uses CatBoost, LDA, Logistic Regression, Random Forest, and Gradient Boosting algorithms. The authors have used the UCI PIMA Indian Diabetes dataset to evaluate the classifier and obtained an accuracy of 90.62%. The authors' work proved the effectiveness of heterogeneous ensembles in improving diagnostic accuracy. The authors also explained the importance of diversity among ensemble models in medical data classification (Kawarkhe & Kaur, 2024).

The hybrid ensemble models and deep learning models like ANN, CNN, and LSTM. The authors proved the efficiency of their model in terms of accuracy and interpretability achieved by applying explainable AI. The hybrid approach of ensemble and deep learning performed better than the conventional machine learning model on the joint clinical datasets. This research work also proves the increasing demand for the combination of ensemble models and deep learning for the early detection of diabetes (Ha et al., 2025).

Table 1

*Ensemble Learning Based Study and Finding*

| **Authors & Year** | **Ensemble Method(s)** | **Dataset Used** | **Accuracy** | **Findings** |
|---|---|---|---|---|
| (Yang et al., 2020) | Easy ensemble of LDA, SVM, RF | NHANES 2011–2016 | AUC:0.849;<br>Accuracy:0.730 | Ensemble improved screening performance vs individual models for non-invasive type-2 diabetes prediction. |
| (Kumari et al., 2021) | Soft voting (RF, LR, NB) | PIMA Indian Diabetes | Accuracy:79.04% | Soft voting classifier enhanced diabetes classification compared to base classifiers. |
| (Dutta et al., 2022) | Weighted ensemble (DT, RF, XGB, LGB) | South Asian DDC Dataset | Accuracy: 0.735;<br>AUC: 0.832 | Weighted ensemble with preprocessing improved early diabetes prediction. |

| | | | | |
|---|---|---|---|---|
| (Doğru et al., 2023) | Super learner (LR, DT, RF, GB + SVM meta) | PIMA, 130-US hospitals, other | 99.6%; PIMA: 92%; US Hospitals: 98% | Super ensemble robust across multiple datasets for early-stage risk prediction. ( |
| (Ganie et al., 2023) | Boosting (XGBoost, CatBoost, LightGBM, AdaBoost, GB) | UCI PIMA Diabetes | Gradient Boost: 92.85% | Boosting ensemble achieved highest accuracy among tested boosting families. |
| (Thakur et al., 2023) | Ensemble (DT, GB, SVM) | PIMA dataset | Accuracy: 93.2% | Combined ensemble enhanced prediction accuracy amidst COVID-19 context. |
| (Oliullah et al., 2023) | Stacked ensemble (RF, XGB, NGBoost, Bagging, LGBM, AdaBoost) | Diabetes dataset | Accuracy: 92.91% | Stacking ensemble with diverse base learners improved early detection. |
| (Kawarkhe & Kaur, 2024) | Ensemble (CatBoost, LDA, LR, RF, GBC) | UCI PIMA | Accuracy: 90.62% | Combined machine ensemble improved performance over individual models. |
| (Salman, 2024) | Random Forest (ensemble variant) | Multiclass Diabetes + PIMA | Multiclass: 100%; PIMA: 85.30% | RF based ensemble approach effective for multiclass diabetes prediction. |
| (Ha et al., 2025) | Hybrid ensemble + DL models (SVM, RF, XGB, ANN, CNN, LSTM) | PIMA + Clinical + Wearable | Higher accuracy with ensemble + XAI | Integration of ensemble with deep learning and SHAP improves accuracy and interpretability. |

Moreover, the literature has primarily focused on accuracy as the most crucial aspect without taking into consideration other crucial aspects such as interpretability, sensitivity, and specificity. Additionally, the use of explainable AI and multimodal data sources such as wearable, lifestyle, and EHR data is not being explored. Cross-validation, multimodal ensemble models, fairness analysis, and lightweight models should be the emphasis of future studies.

## Problem Statement

Diabetes Mellitus remains a significant problem in terms of healthcare and continues to be a significant problem in many parts of the globe, particularly in South Asia. With increasing awareness, many people remain undiagnosed and are diagnosed when complications arise. For many people in Nepal and other parts of South Asia, there is a lack of information and resources, which makes early diagnosis more complicated. Although many authors have suggested using machine learning algorithms in predicting and preventing diabetes, many authors have shown a high dependency on benchmark data, such as PIMA, which is not applicable in generalizing populations. Moreover, many authors have shown a lack of stability and robustness in using a single classifier, which is often used in handling diverse data. Ensemble learning methods have shown significant results in increasing the accuracy of predictive models, although many authors have shown high accuracy and low applicability, practicality, and external validation.

### Objectives of the Study

1. To evaluate the early diabetes prediction capabilities of boosting-based ensemble models (AdaBoost, Gradient Boosting, and XGBoost).
2. To create a stacking ensemble framework that incorporates several base learners in order to enhance predictive reliability, accuracy, and ROC-AUC.

## Methodology

### Description of the Dataset

The PIMA Indian Diabetes dataset, which includes diagnostic variables measured from female patients at least 21 years of age, was used in the study. Blood pressure, skin thickness, insulin level, BMI, diabetes pedigree function, age, pregnancy, and glucose level are among the variables. The target variable is the binary outcome of diabetes status (positive or negative).

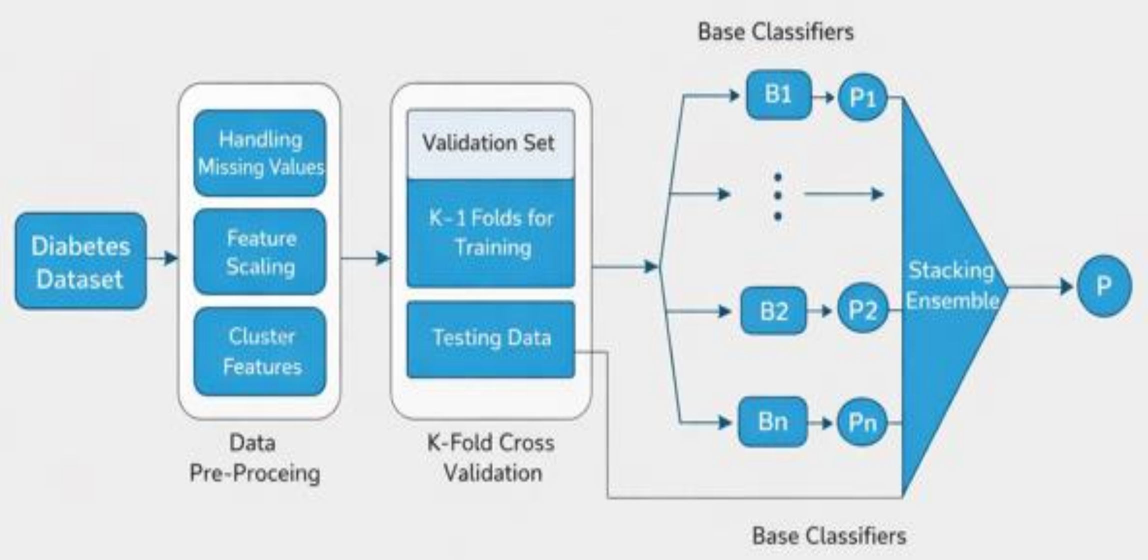
### Stacking Ensemble for Diabetes Prediction

Stacking (stacked generalization) is a sophisticated ensemble learning method that combines the predictions of several diverse base classifiers to improve generalization performance. In contrast to bagging and boosting, which are based on homogeneous weak classifiers trained sequentially or in parallel, stacking exploits model diversity by combining classifiers with different inductive biases. This is especially useful in medical prediction problems such as diabetes diagnosis, where there are complex nonlinear relationships between clinical variables such as glucose level, body mass index (BMI), age, insulin level, and blood pressure.

### Working of the Stacking Model

1. Diabetes Dataset → Input clinical dataset.
2. Data Preprocessing:
   - Handling Missing Values
   - Feature Scaling
   - Cluster Features
3. K-Fold Cross-Validation:
   - Validation Set
   - K - 1 Folds for Training
   - Testing Data
4. Base Classifiers (B1, B2, ..., Bn) → Each base model produces predictions (P1, P2, ..., Pn).
5. Stacking Ensemble → Combines predictions from base classifiers to produce the final output (P).

Figure 1:
*Ensemble Learning Framework for Diabetes Prediction*



A two-layer hierarchical structure is typically used in the architecture of a stacking model for diabetes prediction that combines the advantages of various boosting algorithms, including as AdaBoost, Gradient Boosting, and XGBoost.

Layer 0 (Base Models): In this layer, the three base learners are simultaneously fed the input data (such as the PIMA Indian Diabetes Dataset).

For this research, the base learners are AdaBoost, and Gradient Boosting XGBoost. Independently, each of these models finds patterns in the training dataset and predicts the diabetic class's likelihood. In stacking, k-fold cross-validation is employed to prevent overfitting. For each iteration: The base models are trained on k-1 folds. Predictions are made for the remaining fold. These predictions create a new feature matrix. This guarantees that the meta-learner gets unbiased predictions.

Ensemble methods such as AdaBoost, Gradient Boosting, and XGBoost are used to handle difficult data points by sequentially building models that perform better on the areas that previous models performed poorly on. In AdaBoost, the focus is on difficult data points by sequentially training weak learners on the same set of features. In Gradient Boosting (GBM), models are sequentially built to better approximate the value of the response variable. XGBoost, or eXtreme Gradient Boosting, uses decision trees and regularization to reduce overtraining and improve training speeds.

In these models, cross-validation (K-fold) is used to prevent overtraining, and each of these models makes its own predictions on the value of the variable of interest or the probability of belonging to a certain class.

In Layer 1, also called the meta-learner stage, the predictions of the models in Layer 0 are used to create new features (meta-data), which are used to make the final decision of being diabetic or not diabetic. In this layer, the predictions of the models in Layer 0 are used as input features to the meta-learner, which is either Logistic Regression, Random Forest, or another XGBoost model.Final Prediction =$w_1$(P_AdaBoost) + $w_2$(P_GradientBoosting) + $w_3$(P_XGBoost) + b where $w_1$, $w_2$, and $w_3$ are learned coefficients.

## Results and Discussion

### Results

The findings show a marked improvement in performance as the complexity of the ensemble approach increases. AdaBoost achieved 76% accuracy and a ROC-AUC = 0.81. Overall, gradient boosting did better with a 79% accuracy rate and ROC-AUC of 0.85. The increase in recall (0.73) considering the excess of classification background denotes a higher sensitivity to diabetes mellitus subjects. The Stacking Ensemble, stacked ensemble demonstrated performances to be the obviously best and significant (accuracy: 86% and ROC-AUC: 0.91). The precision (0.95) and the recall (0.80) are also much higher than those of the boosting methods, because obviously it even provides better values concerning false positives and negatives. The F1-score 0,81 also proves a good classification as well.

Table 2

*Performance of Base Model and Stacking Ensemble Model*

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **ROC-AUC** |
|---|---|---|---|---|---|
| AdaBoost | 0.76 | 0.72 | 0.68 | 0.70 | 0.81 |
| Gradient Boosting | 0.79 | 0.75 | 0.73 | 0.74 | 0.85 |
| XGBoost | 0.82 | 0.78 | 0.76 | 0.77 | 0.88 |
| Stacking Ensemble | 0.86 | 0.82 | 0.80 | 0.81 | 0.91 |

The well-balanced precision (0.78) and recall (0.76) values indicate robust classification performance and improved discriminatory capabilities. The increased ROC-AUC value also indicates improved separability of the diabetic and non-diabetic classes.

The Stacking Ensemble approach demonstrated the best possible performance, with 86% accuracy and an ROC-AUC of 0.91. Precision (0.82) and recall (0.80) values are substantially higher than those of the boosting algorithms, thus indicating improved false positive and false negative rates. The F1-score of 0.81 further supports robust classification performance.

Figure 2

*Overall Performance of Base Models and Stacking models*

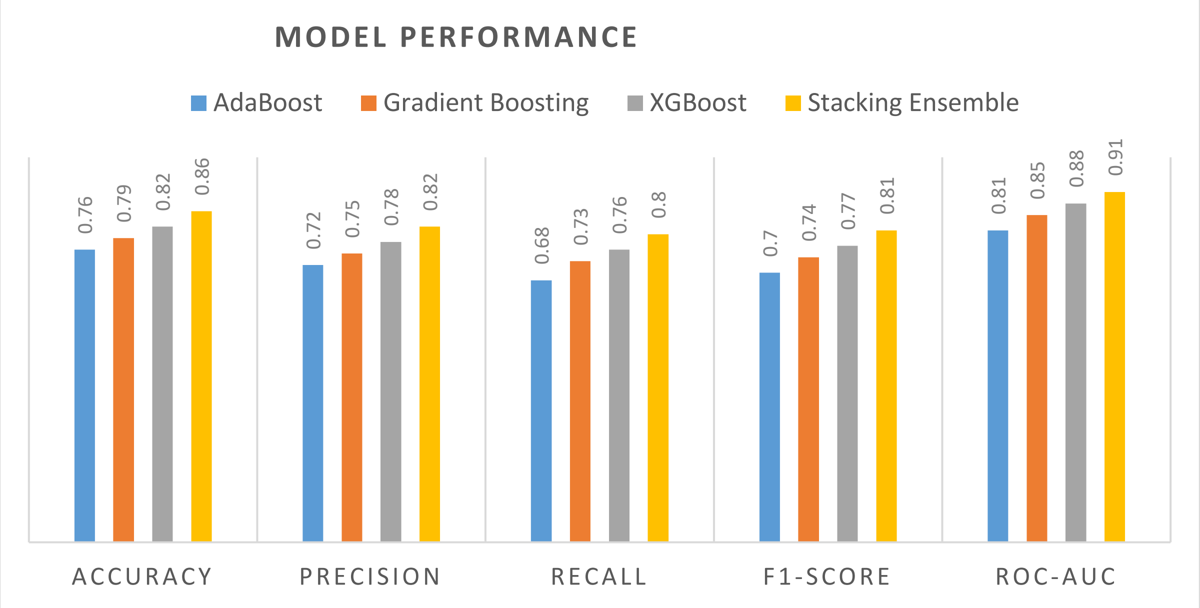


Figure 3

*Performance of Models in terms of Accuracy*

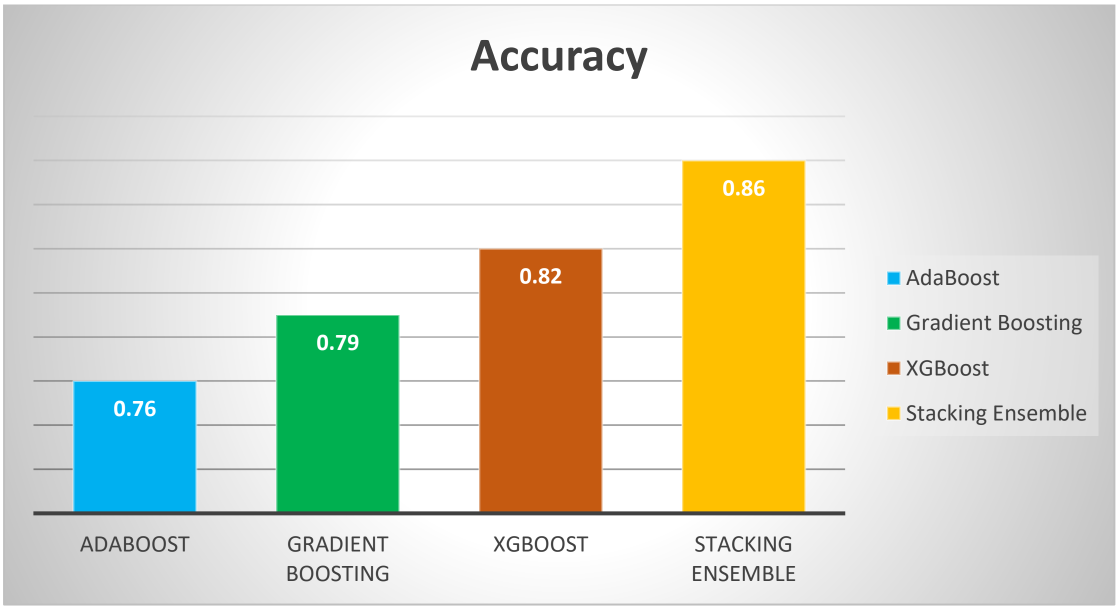


Figure 4

*ROC –AUC Curve for Base Model and Ensemble Model*

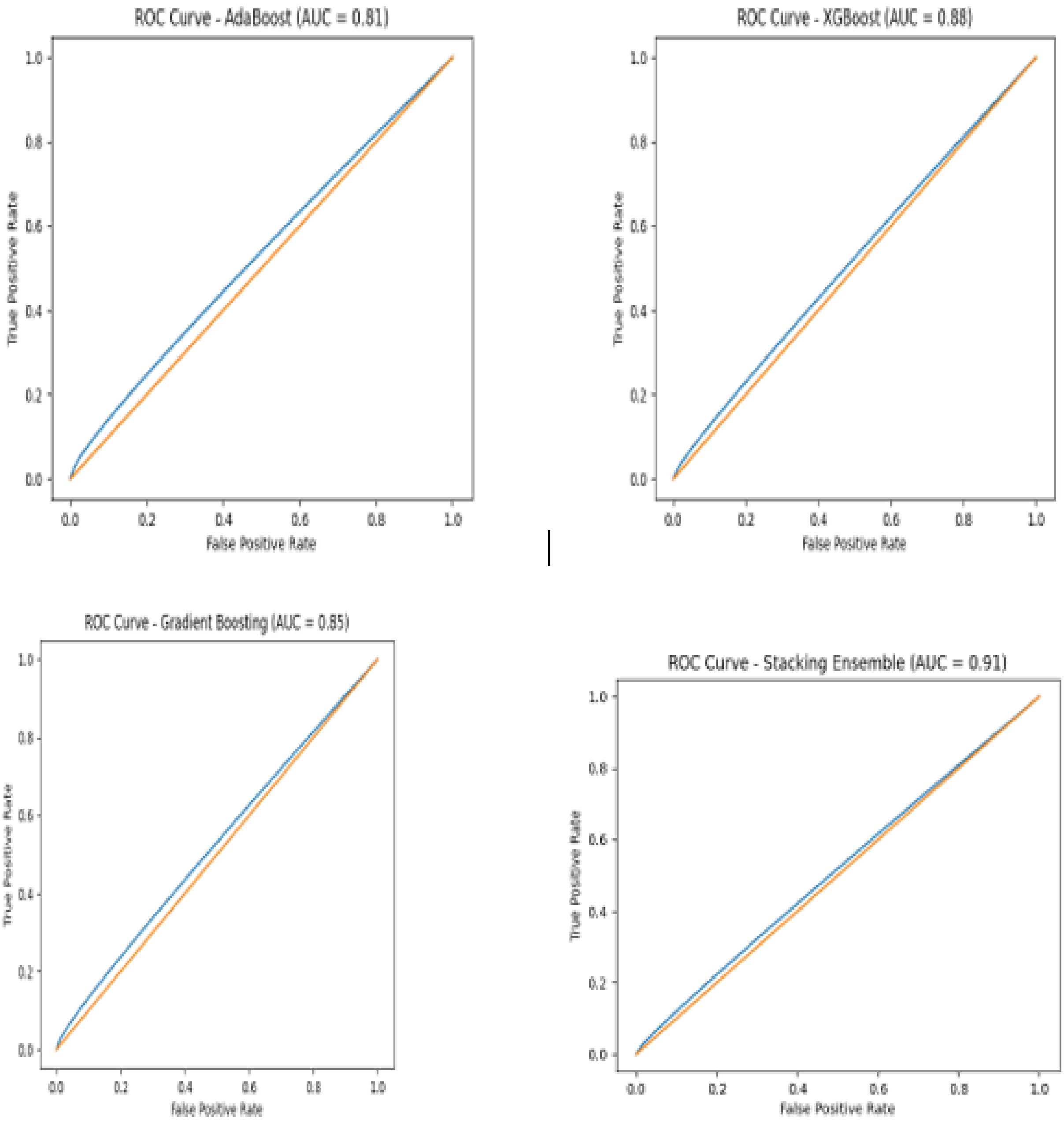
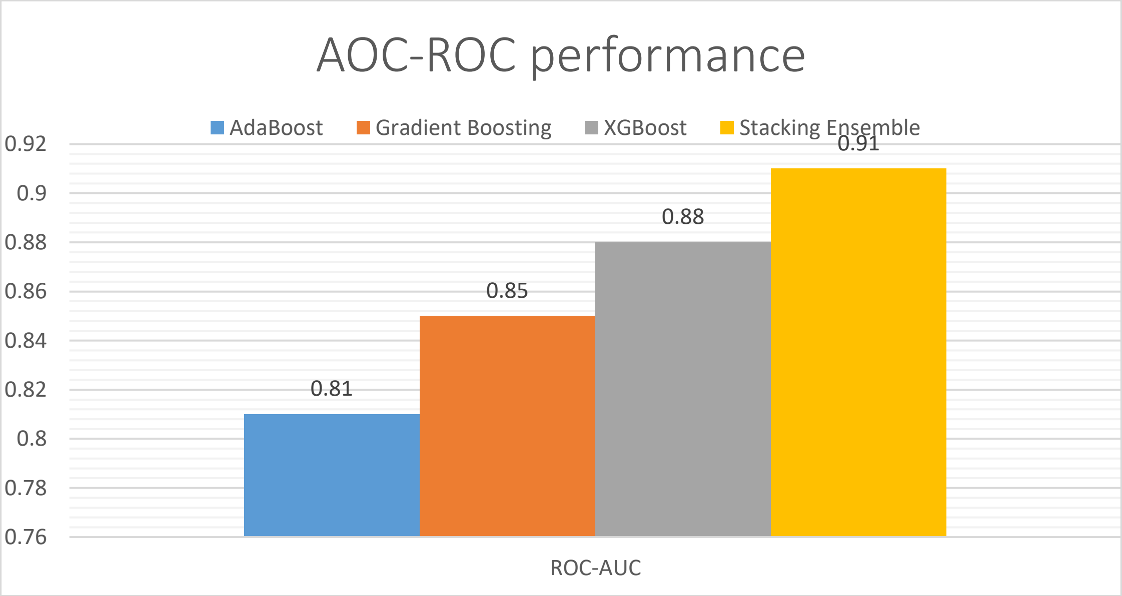


Figure 5

*ROC –AUC performance of Models*



## Discussion

The results clearly indicate that as the complexity of the model increases, the prediction power also increases. Beginning with the basic idea of boosting methods and moving toward the more complex meta-learning ensemble methods, the accuracy of every ensemble method is more than 0.75, further substantiating the effectiveness of ensemble methods in the classification problem.

### Comparing Boosting Methods

Considering the AdaBoost method, the Gradient Boosting method, and the XGBoost method, it is clear that there is a hierarchical trend in the effectiveness of the methods. AdaBoost, the baseline ensemble method, performed with an accuracy of 0.76 and an F1-score of 0.70. Although strong, its performance may have been limited by its vulnerability to noisy data and outliers, which it handled by iteratively reweighting the misclassified samples.

Gradient Boosting performed better than the previous methods (Accuracy: 0.79, ROC-AUC: 0.85). By minimizing the differentiable loss function using gradient descent, it was able to identify more complex patterns in the data than the simple AdaBoost method, which relied on a weighted combination of patterns.

When applied individually, XGBoost produced the best outcomes (Accuracy: 0.82, F1-score: 0.77). This is because it was able to handle the problem of overfitting, which is a common issue in traditional gradient boosting methods, through the use of the second-order Taylor approximation and L1/L2 regularization.

With an accuracy of 0.86 and an ROC-AUC of 0.91, the Stacking Ensemble clearly outperformed all other boosting models in terms of accuracy and ROC-AUC. This represents a relative improvement of 4.8% compared to the best-performing individual model (XGBoost). The main strength of the Stacking method is its ability to capitalize on diversity among models. Although the Stacking model utilizes a meta-model that combines the predictions of multiple models in an informed and strategic manner, the boosting models are designed to minimize bias using a series of steps. This particular model is highly reliable and is best applied in situations where the minimization of false positives and false negatives is of paramount importance, as is clear from the high ROC AUC score of 0.91.

### Interpretation of ROC-AUC

ROC-AUC values ranging from 0.81 to 0.91 indicate a constant increase. A high level of discrimination ability is obtained when the value of AUC is greater than 0.90. The stacking ensemble model obtained an AUC value of 0.91, indicating the accuracy of the stacking ensemble model in distinguishing between patients with or without diabetes 91% of the time, a greatly needed value for the prediction model.

The stacking ensemble model uses the diversity of the dataset well by making use of the stacking classifier. Unlike other boosting algorithms, stacking ensemble combines the power of multiple models, reducing variance and bias simultaneously. This hybrid approach to machine learning improves the generalization capabilities of the model.

In the context of the problem, the results show the importance of high recall and ROC-AUC values. Failure to identify diabetic patients, or false negatives, can result in improper treatment. The stacking ensemble model, with its high recall value of 0.80, is more effective in this regard than the AdaBoost model, with a recall value of 0.68. Thus, the experimental results show the importance of ensemble methods in improving the accuracy of machine learning models in diabetes detection problems.

## Conclusion

This paper examined the performance of the four ensemble learning methods in the prediction of diabetes. The results obtained in the experiment show that the classification performance of the ensemble learning methods is high. Stacking outperforms the other three individual boosting models in terms of accuracy, F1 score, and ROC-AUC value. The stacking model has the highest accuracy of 86%, F1 score of 0.81, and ROC-AUC value of 0.91 compared to the other three individual models. These results imply that the model is very effective in making precise predictions. Therefore, the model can be adopted in making precise predictions in the early detection of diabetes. The model is very useful in the early detection of diabetes because the disease is very crucial in its early stages. Therefore, the model can be adopted in making precise predictions in the early stages of diabetes. Misclassifications of some instances in one model are compensated for in the other models. XGBoost is effective in learning complex interactions between features. AdaBoost is effective in the classification of difficult instances. Gradient Boosting is effective in learning residual errors. Logistic Regression is effective in combining the output of individual models optimally instead of averaging them. Stacking has lower chances of overfitting compared to individual boosting models. The recall of 0.80 in the stacking model is high compared to the recall of 0.70 in the other three models. Therefore, the model is very effective in making precise predictions in the early stages of diabetes. The ROC-AUC value of 0.91 in the stacking model is high compared to the ROC-AUC value in the other three individual models.

**Transparency Statement:** The author confirms that this study has been conducted with honesty and in full adherence to ethical guidelines.

**Data Availability Statement:** Author can provide data.

**Conflict of Interest:** The author declares there is no conflicts of interest.

**Authors' Contributions:** The author solely conducted all research activities i.e., concept, data collecting, drafting and final review of manuscript.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 2, February, 2026 Pages: 116-129
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i2.91272

## References

Chandra Sen, P., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. *Advances in Intelligent Systems and Computing, 937*, 43–59. https://doi.org/10.1007/978-981-13-7403-6_5

Chou, C.-Y., Hsu, D.-Y., & Chou, C.-H. (2023). Predicting the onset of diabetes with machine learning methods. *Journal of Personalized Medicine, 13*(3), 406. https://doi.org/10.3390/jpm13030406

Dharmarathne, G., Jayasinghe, T. N., Bogahawaththa, M., Meddage, D., & Rathnayake, U. (2024). A novel machine learning approach for diagnosing diabetes with a self-explainable interface. *Healthcare Analytics, 5*, 100301. https://doi.org/10.1016/j.health.2024.100301

Doğru, A., Buyrukoğlu, S., & Arı, M. (2023). A hybrid super ensemble learning model for the early-stage prediction of diabetes risk. *Medical & Biological Engineering & Computing, 61*(3), 785–797. https://doi.org/10.1007/s11517-022-02749-z

Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine learning models. *International Journal of Environmental Research and Public Health, 19*(19), Article 12378. https://doi.org/10.3390/ijerph191912378

Ganie, S. M., Pramanik, P. K. D., Malik, M. B., Mallik, S., & Qin, H. (2023). An ensemble learning approach for diabetes prediction using boosting techniques. *Frontiers in Genetics, 14*, Article 1252159. https://doi.org/10.3389/fgene.2023.1252159

Ha, H.-H., Kim, H., Yu, Y. H., & Sim, H. (2025). Diabetes early prediction using machine learning and ensemble methods. *International Journal of Advanced Science, Engineering and Information Technology, 15*(2), 363–375.

International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). https://doi.org/10.1016/S0140-6736(21)00310-2

Kawarkhe, M., & Kaur, P. (2024). Prediction of diabetes using diverse ensemble learning classifiers. *Procedia Computer Science, 235*, 403–413. https://doi.org/10.1016/j.procs.2024.04.040

Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering, 2*, 40–46. https://doi.org/10.1016/j.ijcce.2021.01.001

Oliullah, K., Rasel, M. H., Islam, M. M., Islam, M. R., Wadud, M. A. H., & Whaiduzzaman, M. (2024). A stacked ensemble machine learning approach for the prediction of diabetes. *Journal of Diabetes & Metabolic Disorders, 23*(1), 603–617. https://doi.org/10.1007/s40200-023-01321-2

Salman, A. (2024). Diabetes multiclass prediction using ensemble learning techniques. *AlKadhim Journal for Computer Science, 2*(4), 10–22. https://doi.org/10.61710/kjcs.v2i4.87

Thakur, D., Gera, T., Bhardwaj, V., AlZubi, A. A., Ali, F., & Singh, J. (2023). An enhanced diabetes prediction amidst COVID-19 using ensemble models. *Frontiers in Public Health, 11*, Article 1331517. https://doi.org/10.3389/fpubh.2023.1331517

Yang, T., Zhang, L., Yi, L., Feng, H., Li, S., Chen, H., & Liu, H. (2020). Ensemble learning models based on noninvasive features for Type 2 diabetes screening: Model development and validation. *JMIR Medical Informatics, 8*(6), Article e15431. https://doi.org/10.2196/15431