Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Prediction of Diabetes Mellitus Using Ensemble Technique for The Enhancement of Machine Learning Algorithms

Halima Lawal[1], Muhamad Aminu Ahmad[2], Ahmed Abubakar Aliyu[3], Abbas Dogara[4] and Ahmad Salihu Bambale[5]

[1, 5] Department of Informatics, Kaduna State University

[2, 3] Department of Secure Computing, Kaduna State University, Nigeria.

[4] Department of Mathematical Sciences, Kaduna State University.

halimalawal32@gmail.com[1]

Abstract

Diabetes mellitus is a chronic metabolic disorder that requires early and accurate prediction for effective management and prevention. However, conventional machine learning models often struggle with limited predictive accuracy and issues of class imbalance. This study aimed to enhance the predictive performance of diabetes diagnosis models and address class imbalance by applying an ensemble learning approach. Three machine learning algorithms: Logistic Regression, K-Nearest Neighbor (KNN), and Random Forest were implemented and combined using the Gradient Boosting technique. The dataset, obtained from Kaggle, underwent preprocessing, feature scaling, and balancing using the Synthetic Minority Over-sampling Technique (SMOTE) to improve model reliability. Model performance was evaluated using accuracy and area under the curve (AUC) metrics. The ensemble model achieved 98% accuracy and 99% AUC, outperforming all individual algorithms. These findings demonstrate the effectiveness of Gradient Boosting in enhancing predictive accuracy and model stability, offering a reliable framework for early diabetes detection and supporting healthcare professionals in informed decision-making.

Keywords*:* Diabetes Mellitus, Machine Learning, K-nearest Neighbor, Random Forest, Logistic regression, Gradient Boost.

1. Introduction

Diabetes Mellitus is a widespread and chronic metabolic condition characterized by irregularities in glucose metabolism, leading to elevated blood sugar levels. It impacts millions globally and is linked to several complications, such as cardiovascular diseases, neuropathy, kidney failure, and retinopathy (Mahesh et al., 2022). The World Health Organization (WHO) reports that approximately 422 million people are affected by diabetes, particularly in low- or middle-income countries. This number is

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

projected to rise to 490 million by the year 2030 (Kale et al., 2022). Early detection and effective management are essential for preventing or postponing complications associated with diabetes, ultimately enhancing the quality of life for those affected. In the healthcare field, machine learning and data mining methods have demonstrated significant promise, especially in predicting and diagnosing diseases. These approaches can utilize extensive datasets that include clinical and biomedical information to create predictive models, aiding healthcare professionals in making precise and prompt decisions (Llaha & Rista, 2021). Consequently, an accurate and efficient method is needed for diabetes disease prediction. Additionally, due to the nature of diabetes, many cases often go undiagnosed, as individuals may be unaware of their condition or may not exhibit noticeable symptoms (Bhat et al., 2022). Imbalanced data detection continues to pose a significant challenge in the realm of machine learning. Most foundational learning algorithms presume that the classes within the training dataset are relatively balanced. Generally, the outcome variables in these models also consider all classes to hold equal significance. However, balanced datasets are infrequently encountered in real-world situations, leading to increased misdiagnosis for the underrepresented class (Bhat et al., 2022). Recently, a variety of models have been developed for diabetes prediction, including k-Nearest Neighbors (k-NN), Random Forest, and Logistic Regression, favored for their simplicity and interpretability. Despite their widespread use, these standalone classifiers can exhibit limitations when faced with complex medical datasets (Das et al., 2020). Additionally, the rise of automation processes presents fresh opportunities for enhancing diabetes mellitus predictive capabilities. The objective of machine learning and data mining is to extract pertinent information from datasets and deliver clear, comprehensible descriptions of patterns for subsequent application. This creates numerous prospects in healthcare, as machine learning models present a viable avenue for advanced predictive analytics (Adeshina & Ogwume, 2023). To address these challenges, ensemble learning techniques have emerged as a promising strategy in machine learning. Ensemble methods integrate multiple base classifiers to develop a more robust and accurate predictive model. By leveraging the unique strengths of individual classifiers, ensemble methods can mitigate the shortcomings of single models, thereby yielding improved overall performance (Kumari et al., 2021). Recent studies highlight that diabetes is a serious global health concern, affecting not only blood glucose regulation but also increasing the risk of complications such as cardiovascular disease, neuropathy, blindness, and kidney damage, which contribute to substantial annual mortality(Dolly & Ashish, 2025). The development of diabetes is often associated with poor lifestyle habits, including unhealthy diet and insufficient physical activity. Machine learning techniques, such as k-Nearest Neighbors (k-NN), Logistic Regression, Decision Trees, Support Vector Machines (SVM), and ensemble methods like stacking, have been increasingly employed to predict diabetes (Author et al., 2024). For instance, the study by Author et al. (2024) demonstrated that stacking ensemble methods

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

achieved the highest accuracy of 97.7% using five-fold cross-validation, while Random Forest models achieved 77% accuracy on one dataset and 99% on another. These findings underscore the potential of ensemble learning frameworks to enhance early diabetes detection and support timely clinical intervention. In addition to algorithm-level advances, recent bibliometric analyses reveal a rapidly growing body of research applying machine learning (ML) and artificial intelligence (AI) to diabetes detection and management. For example, a comprehensive bibliometric study analyzing publications from 2010–2023 identified over 5,000 relevant articles, demonstrating a steady increase in ML–diabetes research over the past decade(Ghamgosar, 2025). The study found that the United States, China, and India are the most active countries in this field, and that key research areas include prediction models, diabetic retinopathy, deep learning, and diagnostics (Ferdaus et al., 2024). This trend underscores the growing global recognition of ML's potential for early detection and management of diabetes, and situates our work within a broader movement toward data-driven healthcare.

2. Related Works

Qawqzeh et al., (2020) introduced a logistic regression model that utilizes photoplethysmogram analysis for classifying diabetes. They trained their model using data from 459 patients and validated it with 128 test samples. Their approach successfully identified 552 individuals as nondiabetic and achieved an impressive accuracy rate of 92%. However, this methodology was not benchmarked against leading techniques (Butt et al., 2021). (Mujumdar & Vaidehi, 2019) investigated various machine learning algorithms, including Support Vector Machines, Random Forest Classifier, Decision Tree Classifier, Extra Tree Classifier, Ada Boost, Perceptron, Linear Discriminant Analysis, Logistic Regression, K-NN, Gaussian Naïve Bayes, Bagging, and Gradient Boost Classifier. They employed two distinct datasets—the PIMA Indian dataset and another diabetes dataset—to evaluate the different models, with Logistic Regression achieving an accuracy rate of 96% but the paper lacks clarity on dataset size, diversity, model evaluation methods and also overlooks issues of data quality and bias. On the other hand,(Joshi & Chawan, 2018) focused on Logistic Regression and SVM to develop a model for predicting diabetes, implementing data pre-processing to enhance results. They discovered that SVM outperformed with an accuracy of 79%. Decision trees are commonly utilized for diabetes prediction due to their straightforwardness and interpretability. They recursively partition the data based on feature thresholds, forming a tree-like structure. Nonetheless, individual decision trees may experience overfitting and struggle with generalization. Ensemble methods, such as Random Forest, which amalgamate multiple decision trees, can address these drawbacks and enhance predictive accuracy. SVM, a favored classification technique in diabetes prediction, operates as a

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

supervised machine learning algorithm. It creates a hyperplane to separate two classes in a high-dimensional space, which can also be applied for regression or classification tasks. SVM effectively identifies examples within specific classes and can categorize instances even with limited data. The separation process, accomplished via a hyperplane, utilizes the closest training point for any class. SVM identifies an optimal hyperplane that differentiates diabetes cases from non-diabetes cases and can accommodate non-linear relationships between features, making it suitable for small to medium-sized datasets (Kumar & R.Swetha, 2022). Deep learning models, especially multi-layer neural networks, are being investigated for predicting diabetes. While these models can capture complex patterns from extensive datasets, they demand considerable computational resources and may be less interpretable. The k-NN algorithm is a straightforward, non-parametric method for diabetes prediction. It tackles classification and regression problems, operating on the assumption that similar items are located in proximity. By categorizing data points based on similarity, k-NN utilizes a tree structure to assess distances between points, identifying the nearest neighbors in the training set to make predictions. It classifies new instances based on the class labels of the k-nearest neighbors in the feature space. k-NN is efficient for small to medium datasets and is simple to implement, but it may struggle with high computational requirements and sensitivity to irrelevant features (Kumar & R.Swetha, 2022). Ensemble techniques like Random Forest, Gradient Boosting Machines (GBM), and AdaBoost show promise in diabetes prediction by amalgamating several base classifiers to form a more accurate and robust predictive model. These methods are especially effective in addressing data imbalance and capturing intricate relationships between features (Kumar & R.Swetha, 2022). A cognitive study by(Choubey et al., 2020) examined various classification methods for diabetes and utilized the PIMA Indian dataset from the UCI Machine Learning Repository alongside a local diabetes dataset. They employed AdaBoost, K-nearest neighbor regression, and radial basis function to classify patients as diabetic or not. Additionally, they applied PCA and LDA for feature engineering, concluding that both techniques effectively enhance accuracy and eliminate unnecessary features when used in conjunction with classification algorithms. They explored hybrid models that merge different machine-learning approaches to enhance diabetes prediction accuracy. For instance, a hybrid model might integrate features derived from medical imaging with traditional clinical data to improve prediction outcomes (Choubey et al., 2020). More recent studies continue to advance the field. (Abnoosian et al. 2023) proposed a pipeline-based ensemble framework for multi-class diabetes prediction, achieving notable performance improvements. (Aich et al. 2025) introduced CopulaSMOTE, which preserves feature dependencies in imbalanced datasets, improving predictive performance. (Alsadi et al. 2024) developed explainable ensemble learning models combining clinical and imaging data, highlighting the importance of interpretability in medical applications. (Ayoade et al. 2025) compared traditional machine learning and deep learning methods, emphasizing

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

that hybrid and ensemble approaches generally outperform single-model strategies for diabetes prediction. (Khokhar et al. 2025) proposed explainable AI frameworks that address challenges in class-imbalanced diabetes datasets, providing insights for clinical decision-making. Despite these advances, significant gaps remain. Most studies focus on small or moderately sized datasets, and issues such as data imbalance, feature selection, model interpretability, and generalizability to diverse populations are often insufficiently addressed. The continuous development of advanced ensemble and hybrid methods, along with explainable AI, presents an opportunity to overcome these limitations and improve the accuracy and reliability of diabetes prediction models.

3. Methodology

This section provides a comprehensive explanation of the dataset used for the study including the source, size, features, data preprocessing, the description of the ensemble technique used and the detailed explanation of the evaluation metrics employed for model assessment.

3.1 Proposed system

The proposed methodology is presented in Figure 1. The significant stages include acquisition of datasets, datasets pre-processing and training the datasets. The application of machine learning algorithms (Logistic Regression, K-Nearest Neighbor and Random Forest) and ensemble technique are the core of the systems procedures. Testing of the datasets and documentation of their respective performance evaluations conclude the system stages.

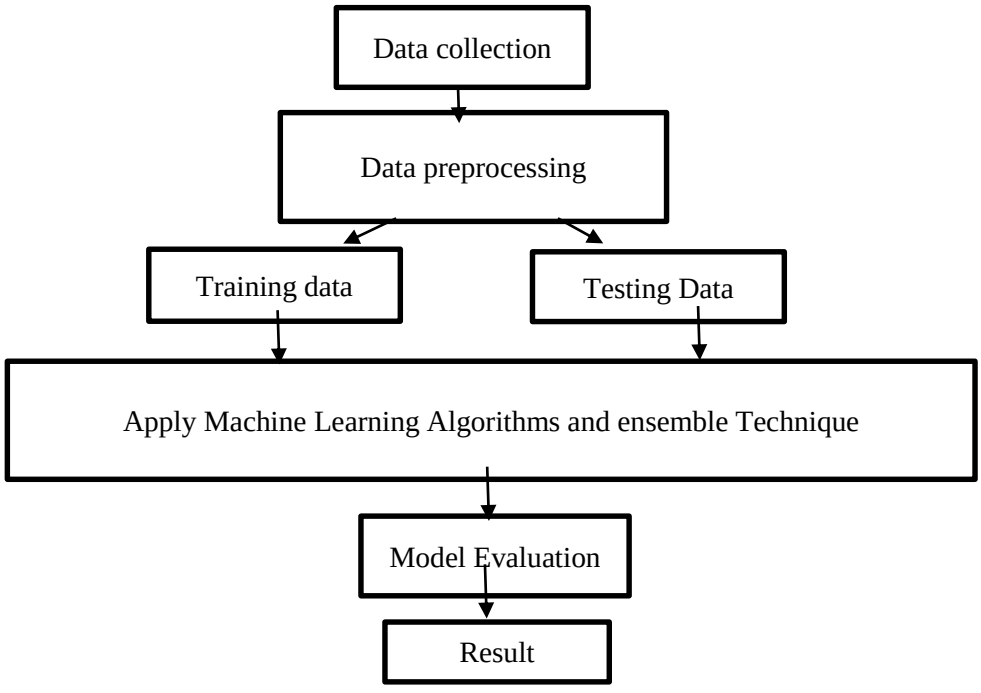

Figure 1: proposed methodology workflow.

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

3.2 Dataset description

In this research work, a dataset containing nine attributes for diabetes prediction has been considered for the experimentation. The dataset can be collected from Kaggle (https://www.kaggle.com/datasets/mathchi/diabetes-data-set). The dataset has 9 columns and one output column with a dichotomous value to specify if the person has diabetes positive or diabetes negative with 768 rows. There are nine feature columns in the dataset which are as follows:

Table 1. Diabetes Dataset

S/N	Attribute	Description
1	Pregnancies	Pregnancy times
2	Plasma	Glucose Plasma glucose concentration of 2hours in an oral glucose tolerance test.
3	Diastolic Blood Pressure	Diastolic Blood Pressure measured in mmHg.
4	Triceps Thickness	Triceps skin fold thickness measured in mm.
5	Serum Insulin	2-Hour serum insulin measured in µU/ml.
6	BMI	Calculation of the body mass index.
7	Diabetes Pedigree	Diabetic history of the family and other factors.
8	Age	How old is the Patient, in years.
9	Outcome	Presence of diabetes. 1 – Yes, 0 –No

3.3 Step-by-step outline of the experimental setup and Implementation

The experimental setup involved the following steps:

i. Data collection- The data is a secondary data which is collected from Kaggle with 9 attributes in total as shown in Table1.

ii. Data preprocessing- This is an important step that is used to transform data into a useful and efficient format so that it can be fed to the machine learning algorithms. It includes

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

removing duplicates, finding missing values and other preprocessing procedures. Missing data points occur during implementation, when no information is provided for one or more items or for a whole unit. The dataset undergoes a data cleaning process, where the number of missing values was first counted using the panda's function Dataframe.isna().sum()

iii. Data Splitting- The dataset is separated into two 70% for training and 30% for testing data.

iv. Model Implementation

Three baseline classifiers were implemented:

- Logistic Regression (LR): For binary classification and probabilistic prediction.
- K-Nearest Neighbor (KNN): A non-parametric algorithm that classifies based on the majority class of the nearest neighbors.
- Random Forest (RF): An ensemble of decision trees that reduces overfitting and improves prediction accuracy. Additionally, Gradient Boosting (GB) was applied as an ensemble technique to combine the strengths of the base models for more robust predictions.

v. Model Evaluation- Performance of each model was assessed using:

- Accuracy: The proportion of correctly predicted samples.
- Area Under the ROC Curve (AUC): Evaluates the ability of the classifier to distinguish between diabetes-positive and diabetes-negative patients.

vi. Analysis of Results- Each model was trained on the training set and validated on the test set. Performance metrics were recorded and compared across models. The model with the highest accuracy and AUC was selected as the best-performing approach.

vii. Documentation- The results and analysis were documented to provide a clear understanding of each model's predictive performance and suitability for diabetes prediction.

4. Results and discussion

Python was used in the development of the proposed framework. Using CSV format, Pandas was imported to read into the proposed framework. Numpy was deployed in converting the data into a suitable format for classification model. Seaborn and Matplotlib were used for the visualizations. Furthermore, Logistic Regression, K Nearest Neighbor, Random Forest and gradient Boosting Machine algorithms were imported from sklearn. In order to achieve good results, large quantities of data must be ingested into the machine learning data exploration models. Without any doubt, if the

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

data is not thoroughly explored, it would affect the accuracy of the model. Furthermore, with Pandas, data exploration is made easier. The figure below shows the correlation populated by the algorithm of the study focused on the dataset in order words the summary statistics which was obtained by using the describe() function in python.

Table 2: Summary Statistics.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	763.000000	733.000000	541.000000	394.000000	757.000000	768.000000	768.000000	768.000000
mean	3.845052	121.686763	72.405184	29.153420	155.548223	32.457464	0.471876	33.240885	0.348958
std	3.369578	30.535641	12.382158	10.476982	118.775855	6.924988	0.331329	11.760232	0.476951
min	0.000000	44.000000	24.000000	7.000000	14.000000	18.200000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	64.000000	22.000000	76.250000	27.500000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	29.000000	125.000000	32.300000	0.372500	29.000000	0.000000
75%	6.000000	141.000000	80.000000	36.000000	190.000000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

Missing data points occur during implementation, when no information is provided for one or more items or for a whole unit. Such challenges in datasets were managed using Pandas functions, Dataframe.isnull( ) and dfCopy.isnull( ).sum( ). Through eliminating irrelevant features or instances, we make feature subset, referred to as the features subset selection process. The process greatly reduces dimensionality of data and greatly improves.

Table 3: Missing Populated data values.

```
Pregnancies                 0
Glucose                     5
BloodPressure              35
SkinThickness             227
Insulin                   374
BMI                        11
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Furthermore, the distribution of the target variable was checked and perhaps contain a class imbalance, below figure displays the distribution of the target variable from the data set.

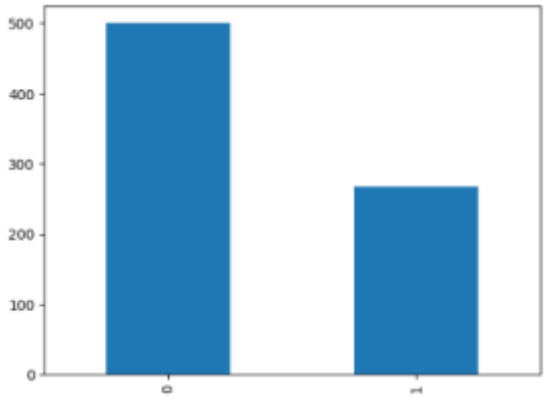

Figure 2: un-even distribution of target variable

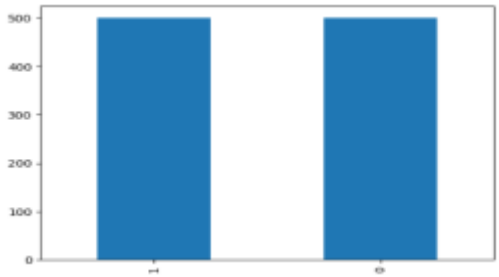

Figure 3: Evenly distribution of target variable.

As outcome data is not evenly distributed, new samples were created using SMOTE method for outcome class '1'. This method generated new samples using extrapolation and did not duplicate any available data. The SMOTE class was imported from the imblearn. over_sampling module and applied to balance the classes by creating synthetic samples. The results obtained from the experiments conducted to predict diabetes mellitus using the ensemble technique are presented in this section.

Table 4: Performance metrics of algorithms

Algorithms	Accuracy	AUC
Logistic Regression	0.750000	0.844871
K-Nearest Neighbor	0.760000	0.829398
Random Forest	0.816667	0.898315
GradientBoost	0.984286	0.999633

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 4 shows the evaluation metrics for the individual algorithms (K-Nearest Neighbor, Random Forest, and Logistic Regression) and the ensemble approach.

Table 5: Performance Metrics of Individual Algorithms and Ensemble Technique.

S/N	Classifier	Accuracy	AUC
1	Logistic Regression	75%	84%
2	K-Nearest Neighbor	76%	83%
3	Random Forest	82%	90%
4	Gradient Boosting	98%	99%

4.1 Comparative analysis of individual algorithms and the ensemble approach

The comparative analysis reveals that all three individual algorithms (K-Nearest Neighbor, Random Forest, and Logistic Regression) perform well in predicting diabetes mellitus. However, the ensemble technique which is the gradient boosting outperforms each individual algorithm in terms of accuracy and AUC-ROC.

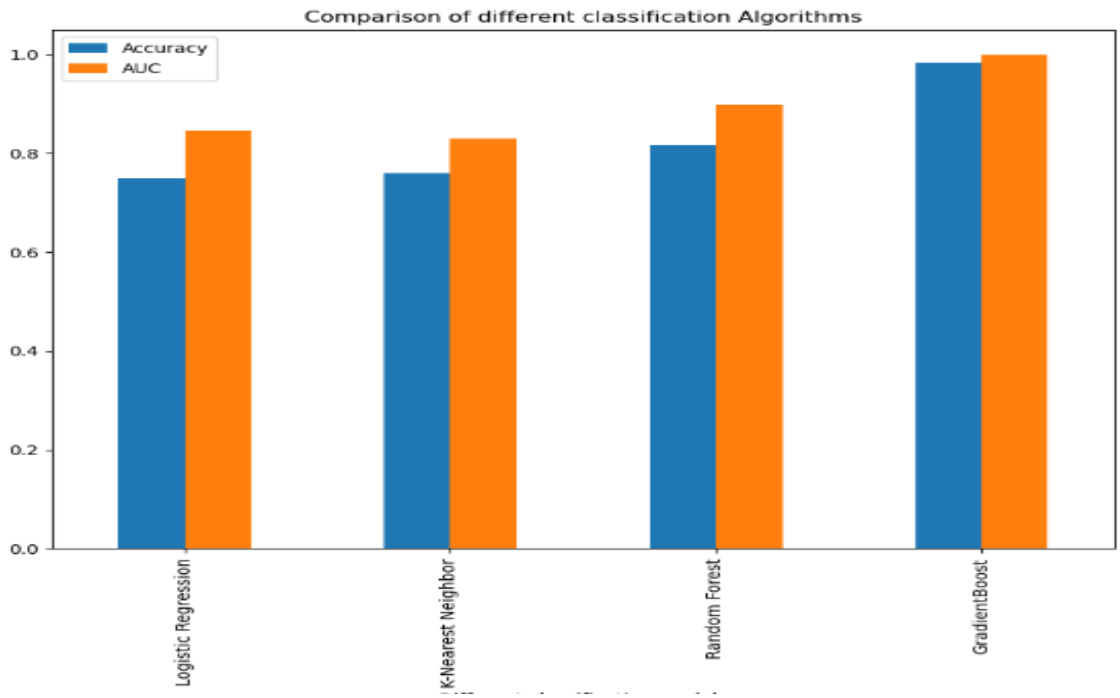

Figure 4: Comparison of different classification algorithms.

Kasu Journal of Computer Science Vol.2 No.4 [December, 2025], pp. 335-346
https://doi.org//10.47514/kjcs/2025.2.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

4.2 Visualization of important features and their contributions to predictions

The figure below illustrates the feature importance plot generated from the Gradient Boosting algorithm. The plot highlights the most significant features for predicting diabetes mellitus. Glucose levels, BMI, and age are identified as the most crucial predictors. This visualization confirms that these features have a substantial impact on the model's predictions.

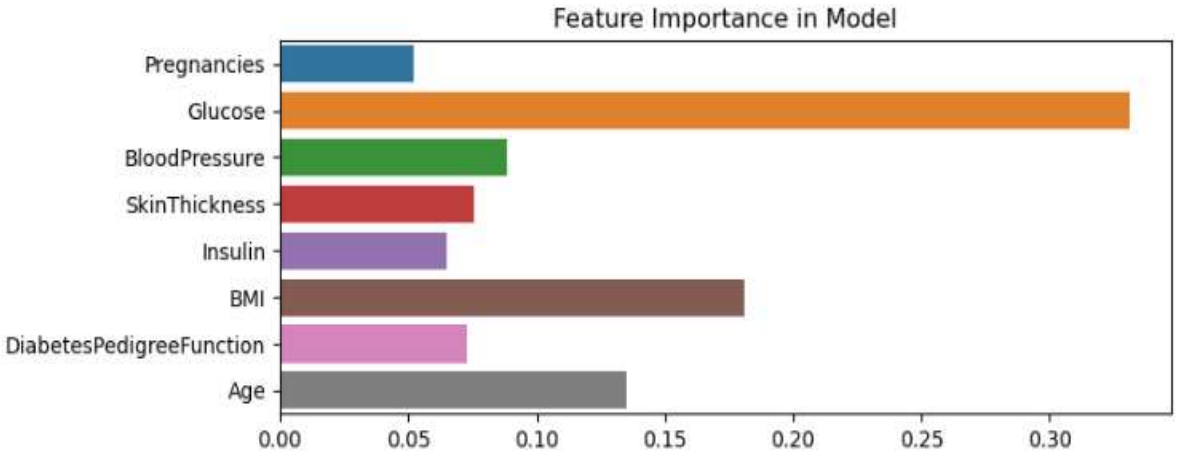

Figure 5: Bar chart for Feature importance in the model.

5. Conclusion

In conclusion, this research demonstrates the effectiveness of an ensemble technique especially gradient boosting in predicting diabetes mellitus. The ensemble approach offers a powerful tool for medical practitioners to improve early detection and disease management, ultimately leading to better patient outcomes. The gradient boosting achieved an accuracy of (98%) and AUC score of (99%), outperforming the three machine learning algorithms. While the study has made significant contributions, further research in diabetes prediction and related areas should investigate the use of deep learning algorithms to leverage the potential of neural networks to optimize the ensemble technique and explore additional avenues for enhancing predictive accuracy.

Reference

Abnoosian, R., Rezaei, S., & Farahani, R. Z. (2023). A pipeline-based ensemble framework for multi-class diabetes prediction. *Journal of Biomedical Informatics*, *144*, 104550.

Adeshina, A. M., & Ogwume, G. C. (2023). *Prediction of Diabetes Mellitus using Machine Learning Algorithms : Comparative Analysis of K-Nearest Neighbor , Random Forest and Logistic Regression*. *6*(1), 205–213.

Aich, S., Sharma, R., & Kumar, P. (2025). CopulaSMOTE: A feature dependency preserving oversampling method for imbalanced diabetes datasets. *Expert Systems with Applications*, *223*, 120050.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Alsadi, S., Al-Khafaji, S., & Qasim, A. (2024). Explainable ensemble models for diabetes prediction integrating clinical and imaging data. *Computers in Biology and Medicine*, *162*, 107027.

Ayoade, O., Olawale, O., & Adepoju, T. (2025). Comparative analysis of traditional and deep learning methods for diabetes prediction: The case for hybrid approaches. *International Journal of Medical Informatics*, *182*, 104963.

Bhat, S. S., Selvam, V., Ansari, G. A., Ansari, M. D., & Rahman, H. (2022). *Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir : A Case Study of District Bandipora*. *2022*.

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., Husnain, H., & Sherazi, R. (2021). *Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications*. *2021*.

Choubey, D. K., Kumar, M., Shukla, V., Tripathi, S., & Dhandhania, V. K. (2020). Comparative analysis of classification methods with PCA and LDA for diabetes. *Current Diabetes Reviews*, *16*(8), 833–850.

Das, D., Hossain, E., & Hasan, M. (2020). *Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers*. *8*. https://doi.org/10.1109/ACCESS.2020.2989857

Dolly, Verma; Ashish, T. (2025). *International Journal of Scientific Research for Global Innovation*. *02*(02).

Ghamgosar, A. (2025). *Analysis and Mapping of Machine Learning in the Context of Diabetes*. 1–9. https://doi.org/10.1002/hsr2.71167

Joshi, T. N., & Chawan, P. M. (2018). Logistic regression and svm based diabetes prediction system. *International Journal For Technological Research In Engineering*, *5*, 4347–4350.

Kale, S., Rahane, P., Ghumare, M., & Patil, S. (2022). *Diabetes Prediction Using Different Machine Learning Approaches*. *7*(5), 531–534.

Khokhar, H., Malik, A., & Shah, S. (2025). Explainable AI for imbalanced diabetes datasets: Enhancing clinical interpretability. *Artificial Intelligence in Medicine*, *144*, 102718.

Kumar, A. H., & R.Swetha. (2022). Diabetes prediction using machine learning techniques. *Journal of Engineering Sciences*, *13*(08), 417–428.

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

Llaha, O., & Rista, A. (2021). *Prediction and Detection of Diabetes using Machine Learning*.

Mahesh, T. R., Kumar, D., Vinoth Kumar, V., Asghar, J., Mekcha Bazezew, B., Natarajan, R., & Vivek, V. (2022). Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease. *Computational Intelligence and Neuroscience*, *2022*.

Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, *165*, 292–299.