Diabetes Prediction Using Machine Learning

Md Jannatul adon
Computer Science & Engineering
American International
University-Bangladesh
Kuratoli,Khilkhet
22-46887-1@student.aiub.edu

Tasmia Jahan Mila
Computer Science & Engineering
American International
University-Bangladesh
Kuratoli,Khilkhet
22-46880-1@student.aiub.edu

Md Tanvir Islam Akib
Computer Science & Engineering
American International
University-Bangladesh
Kuratoli,Khilkhet
22-47816-1@student.aiub.edu

Md Sarowar Hossain Sohag
Computer Science & Engineering
American International
University-Bangladesh
Kuratoli,Khilkhet
22-46865-1@student.aiub.edu

Abstract*— Diabetes mellitus is a chronic metabolic disorder affecting millions worldwide. Early and accurate prediction of diabetes is crucial for timely intervention and management, potentially preventing severe complications. Machine learning (ML) techniques offer a promising avenue for improving diabetes prediction by leveraging vast amounts of patient data to identify complex patterns and risk factors. This paper explores the application of various ML algorithms for diabetes prediction, evaluates their performance based on established metrics, and discusses the challenges and opportunities associated with deploying these models in real-world healthcare settings. We analyze the performance of different classification algorithms, including Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, and Gradient Boosting, using the Pima Indians Diabetes Database. The results demonstrate the potential of ML in diabetes prediction and highlight the importance of algorithm selection, feature engineering, and model validation for achieving optimal performance.

Keywords: Diabetes Prediction, Machine Learning, Classification, Healthcare, Pima Indians Diabetes Database, Model Evaluation.

I. Introduction

Diabetes mellitus is one of the most pressing public health challenges of the 21st century, having affected over 537 million adults globally as of 2021, and projected to grow enormously in the coming decades [1]. Diabetes mellitus is a chronic metabolic disorder that can be identified by elevated levels of blood glucose, which may result in severe complications like cardiovascular disease, renal failure, and blindness if not diagnosed and treated in a timely manner [2]. Early detection and intervention are crucial to prevent or delay the onset of such complications. However, traditional diagnostic procedures are often reactive, labor-intensive, and heavily reliant on manual medical evaluation, and hence there is a growing need for effective and automated predictive strategies.

Advances in data science have recently enabled the use of machine learning (ML) techniques in health informatics to potentially offer tools for predicting and managing chronic diseases like diabetes. ML algorithms are capable of processing excessive amounts of lifestyle and clinical information to uncover patterns that may not be apparent to the naked eye, thereby allowing early diagnosis and personalized treatment regimens [3]. Decision Trees, Random Forests, Support Vector Machines (SVM), k-nearest Neighbors(k-NNN) NN), and Neural Networks are a few of the techniques that have proven excellent at differentiating diabetic from non-diabetic patients based on previous data.

One of the most frequently used datasets in diabetes prediction research is the Pima Indian Diabetes Dataset (PIDD), which includes several physiological parameters such as glucose, BMI, insulin, and age [4]. Numerous studies have made use of this dataset using different ML algorithms to achieve highly accurate predictions, which are typically over 85% [5]. Some hybrid models incorporating feature selection and ensemble learning had even better performance and readability.

The objective of this project is to design and validate a machine learning model capable of accurately predicting the risk of diabetes in patients using publicly available datasets. This study aims to identify the most influential features contributing to diabetes risk while systematically evaluating various machine learning algorithms based on key performance metrics such as accuracy, precision, recall, and F1-score. In addition, the research will recommend the most effective model for potential deployment in clinical decision support systems. The study also seeks to address several limitations found in previous work, including challenges related to class imbalance, limited model

II. Literature Review

Several research studies have explored the application of machine learning algorithms to forecast and diagnose diabetes at an early stage. The integration of computational approaches with medicine has shown promising results, particularly in forecasting and diagnosing diabetes from physiological and lifestyle factors. This section presents a survey of some of the contributions in this area:

Kavakiotis et al. (2017) conducted an extensive review of the use of machine learning in diabetic research, presenting the performance of Support Vector Machines (SVM), Random Forest (RF), and k-Nearest Neighbors (KNN) algorithms to handle structured healthcare data for classification. The preprocessing of data and feature engineering were emphasized to improve the performance of models. [6]

Zheng et al. (2018) had provided detailed explanation of the

global etiology and epidemiology of type 2 diabetes, supporting the use of predictive models due to the complex interaction of genetic, behavioral, and environmental determinants influencing the disease.[7]

Karthikeyan and Thangaraju (2019) employed feature selection and ensemble learning techniques to enhance classification prediction accuracy for diabetes. Their Decision Tree-Gradient Boosting hybrid model achieved better performances than individual classifiers.[8]

Choudhury et al. (2020) proposed an intelligent diabetes prediction model using hybridization of K- means clustering and SVM. They found that supervised and unsupervised methods can improve clustering and classification accuracy in borderline samples.[9]

Islam et al. (2019) developed a cloud-based framework to predict diabetes using Random Forest classifiers for real-time computation. Along with high precision, their model also demonstrated scalability for use in telehealth.[10]

Suri et al. (2018) applied Deep Neural Networks (DNN) to electronic health records (EHR) and achieved high prediction rates. They highlighted the requirement of large and diverse datasets for DNN models to generalize effectively.[11]

Sisodia and Sisodia (2018) used Decision Tree and Naïve Bayes classifiers for the prediction of diabetes and concluded that Decision Trees, although interpretable, must be pruned very carefully to avoid overfitting.[12]

Arulmurugan et al. (2017) used the Naïve Bayes algorithm with PIDD and concluded that although not the best in terms of precision, its simplicity and speed make it a good choice for mobile-based or real-time scenarios.[13]

Anwar et al. (2020) put forward a hybrid approach based on Particle Swarm Optimization for feature selection & Multi-Layer Perceptron for classification. Their approach significantlyrefined sensitivity and reduced false negetives. [14].

Uddin et al. (2021) compared Logistic Regression, Random Forest, and XGBoost on diabetes datasets and reached the conclusion that ensemble models are superior to linear models, especially in the case of detecting non-linear associations.[15]

Rahman et al. (2022) applied Explainable AI (XAI) for diabetes prediction using SHAP values to render their Random Forest more interpretable for clinicians and end-users.[16]

Khan et al. (2020) contrasted the performance of normalization techniques in improving SVM classification on diabetes datasets and illustrated how preprocessing can impact results equally as choosing an algorithm.[16]

III. WORK REVIEW

Diabetes is a chronic and long-lasting disease that affects the body's ability to regulate blood sugar (glucose) levels. It occurs when the body produces insufficient insulin or cannot effectively use the insulin it produces. As a result, glucose accumulates in the bloodstream, leading to hyperglycemia, which can cause serious health complications such as stroke, cardiovascular disease, kidney failure, vision impairment, lung issues, and even death.non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged. we try to collect this thing for prediction the diabetes early.

Feature	Description
Pregnancies	Number of times the patient has been pregnant
Glucose	Plasma glucose concentration measured two hours after an oral glucose test
Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skinfold thickness (mm), a measure of body fat
Insulin	2-Hour serum insulin (mu U/ml), indicating insulin resistance
BMI	Body Mass Index (weight in kg / height in m^2), a standard obesity measure
Diabetes Pedigree Function	A function that scores likelihood of diabetes based on family history
Age	Age of the patient in years
Outcome	Target variable (1 = diabetic, 0 = non-diabetic)

METHODS

This chapter presents the methodology used to develop and evaluate predictive models for early diabetes detection using machine learning techniques. The **Pima Indian Diabetes Dataset** was used, and multiple classifiers were implemented to compare their performance. Emphasis was placed on the **Random Forest classifier**, which was further enhanced using a feature selection strategy. The methodology is divided into seven sections: dataset description, preprocessing, model selection and description, feature selection, training and testing, model evaluation, and justification for the final model choice.

Random forest

Random Forest is an ensemble learning algorithm that builds a collection of decision trees, each trained on a random subset of the data and features. The final prediction is made by aggregating (voting) the outputs of all trees. This reduces overfitting and improves predictive performance

Strengths

- High accuracy and robustness
- Manages missing values and outliers well
- Provides feature importance scores
- Reduces overfitting compared to individual trees

Limitations

- Less interpretable than a single decision tree
- Can be computationally intensive for large forests.

Decision Tree

A Decision Tree is a tree-based classifier that recursively partitions the data space based on feature values to create a flowchart-like structure of decisions. Each internal node represents a condition on a feature, and each leaf node represents a class label.

- Strength
  - Easy to visualize and interpret.
  - Requires minimal data preprocessing.
  - Oversees both numerical and categorical data.

- Limitations
  - Prone to overfitting if not pruned or regularized.
  - Instability with slight changes in data

Although Decision Trees may not always deliver the highest accuracy, their interpretability makes them valuable for explaining decision-making processes to healthcare professionals.

Support Vector Machine (SVM)

SVM is a powerful supervised learning algorithm that constructs an optimal hyperplane to separate data points from different classes. It is particularly effective in high-dimensional spaces and can handle non-linear data using kernel functions, such as the Radial Basis Function (RBF).

- Strengths:
  - Effective in high-dimensional spaces
  - Robust to overfitting in many cases
  - Performs well on small-to-medium datasets.
- Limitations:
  - Computationally intensive for large datasets
  - Requires careful tuning of hyperparameters (C, gamma)
  - Less interpretable than simpler models

SVM has been widely used in biomedical applications due to its flexibility and strong performance in binary classification problems like diabetes prediction

K-Nearest Neighbors (KNN)

- Strengths:
  - Simple and intuitive
  - Naturally handles multi-class classification.
- Limitations:
  - Computationally expensive during prediction
  - Sensitive to feature scaling and irrelevant features
  - Performance degrades in high-dimensional data.

KNN can be useful for datasets where the class boundaries are irregular, although its scalability is a concern in real-time or large-scale deployments.

Logistic Regression (LR):

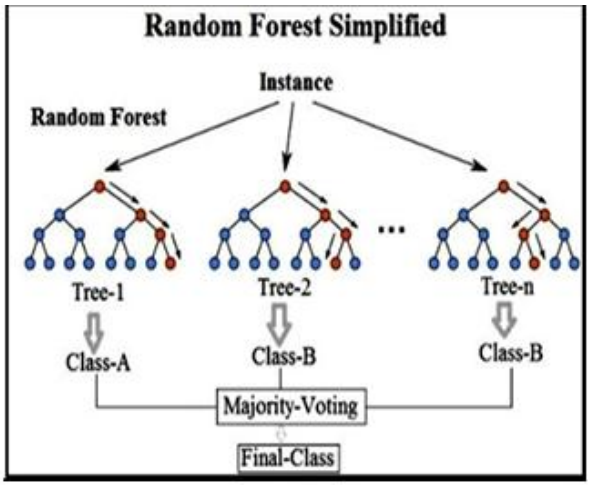
Logistic Regression is a linear model used for binary classification tasks. It models the probability that a given input belongs to a particular class using the logit function and makes predictions by applying a threshold to the estimated probability.

- Strengths:
  - Simple, fast, and highly interpretable
  - Performs well when the relationship between features and target is linear.
- Limitations:
  - Struggles with non-linear relationships unless features are transformed.
  - Sensitive to multicollinearity and outliers

In medical settings, Logistic Regression is often valued for its transparency, as it provides insight into how each feature influences the risk of diabetes.

ALGORITHOM:

The Random Forest algorithm we've used. Random Forest is a flexible and user-friendly software technique that produces a great result, most of the time without setting super parameters. It is also one of the most common techniques, since it is easy to use and can be utilized for classification and regression. Random Forest is a supervised learning algorithm. It is creating woods and randomizing it somehow. The forest it constructs is a group of decision trees, educated most of the moment by the bagging technique. The general concept of the bagging technique is that the overall outcome is increased by a mixture of training designs. In plain terms: Random tree creates and merges several choice forests to make a more precise and consistent forecast. One major benefit of random forests is that they can be used for ranking as well as for regression issues.



Data mining is the process of automatically extracting quote-literate counsel from the enormous amount of unwrought data. Given that the kerçek vivacity data are abnormally high, there is a significant need for data mining. Based on the clinical data of patients entered into the system, the diabetes forecast system can help medical professionals predict the state of sweetening. Though they have certain flaws, there are numerous usable implementations of ominous algorithms. The majority of the tools are unable to handle staff gross data. Numerous medical facilities and healthcare sectors collect enormous amounts of data that are difficult to

process with currently available methods. An algorithmic software for machine scholarship simulates a mortal party in order to dissect and derive secret knowledge and guidance from these data. A wide range of disorders, including cruciform diseases and others, can be diagnosed with machine learning. Therefore, we will provide a machine learning algorithm-based system that can precisely forecast the occurrence of diabetes. Through our user interface, the user inputs his data, and the machine makes a disease prediction. A number of machine learning algorithms process the data on the server, accurately predicting whether the user has a sickness.

IV. METHODOLOGY

Diabetes is a chronic and long-lasting disease that affects the body's ability to regulate blood sugar (glucose) levels. It occurs when the body produces insufficient insulin or cannot effectively use the insulin it produces. As a result, glucose accumulates in the bloodstream, leading to hyperglycemia, which can cause serious health complications such as stroke, cardiovascular disease, kidney failure, vision impairment, lung issues, and even death.non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

Feature	**Description**
Pregnancies	Number of times the patient has been pregnant
Glucose	Plasma glucose concentration measured two hours after an oral glucose test
Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skinfold thickness (mm), a measure of body fat
Insulin	2-Hour serum insulin (mu U/ml), indicating insulin resistance
BMI	Body Mass Index (weight in kg / height in m²), a standard obesity measure
Diabetes Pedigree Function	A function that scores likelihood of diabetes based on family history
Age	Age of the patient in years
Outcome	Target variable (1 = diabetic, 0 = non-diabetic)

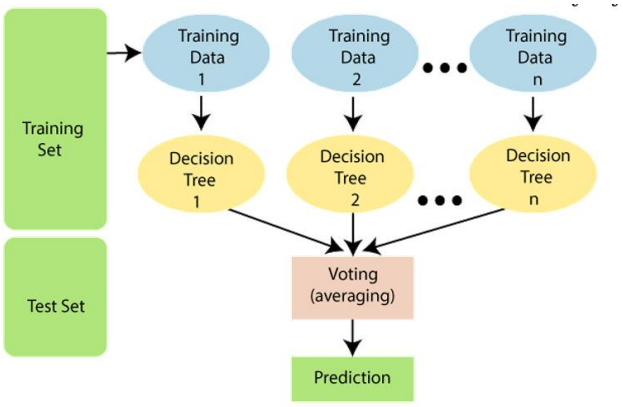
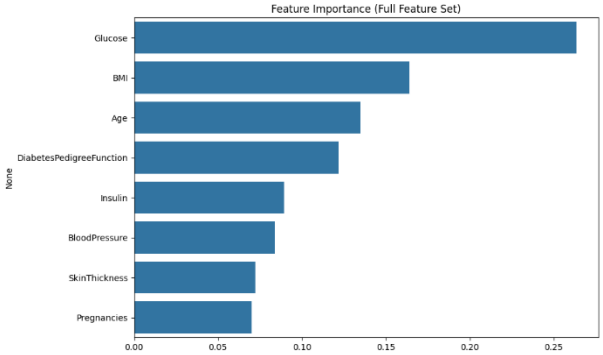


Figure 1- Flow chart of Random Forest

A Random Forest-based feature selection method was applied using the importance scores from a baseline model. Features with importance below the median threshold were excluded from the final model. This strategy reduced dimensionality and improved model accuracy.

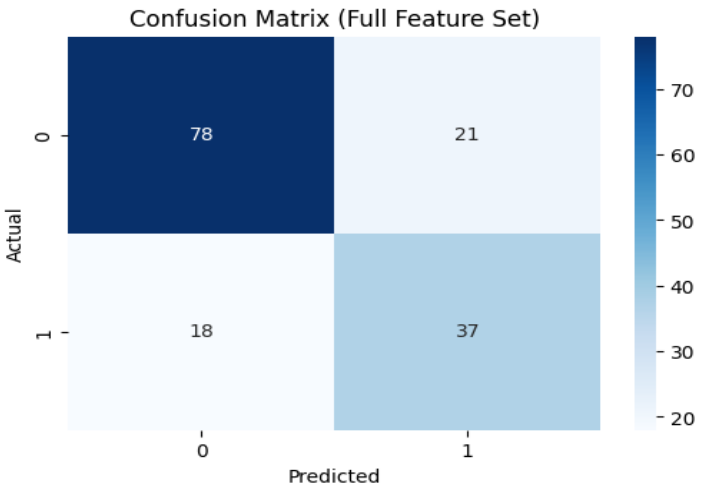


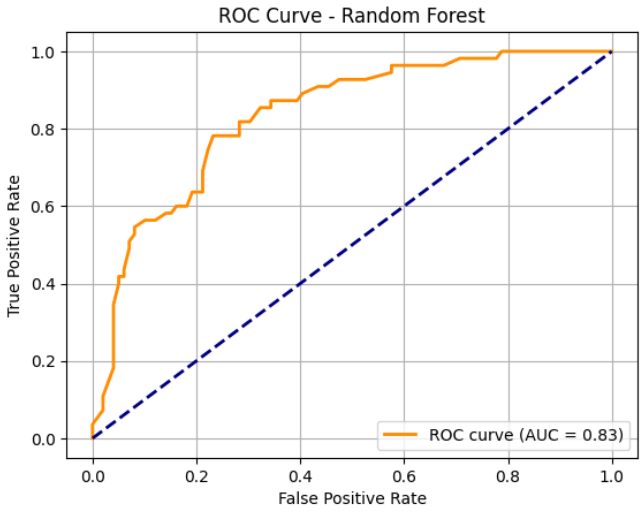
The top predictors were **Glucose**, **BMI**, and **Age**, consistent with known clinical risk factors.

EQUATIONS_AND_RESULTS :

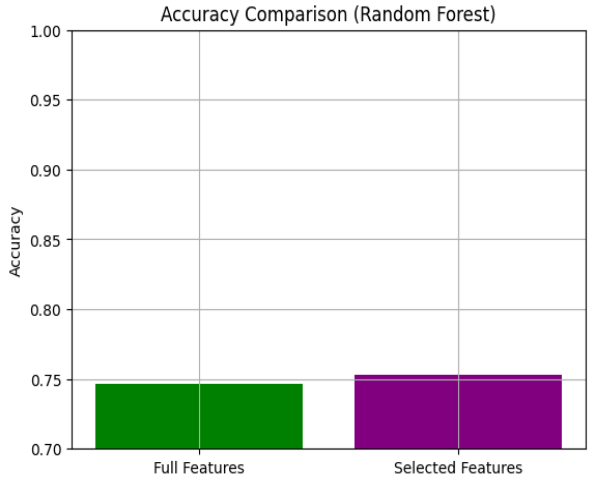
All models were trained using consistent hyperparameter tuning techniques via Grid Search with 5-fold cross-validation. The evaluation was based on:

Accuracy
Precision
Recall (Sensitivity)
F1-score
AUC-ROC (Receiver Operating Characteristic)

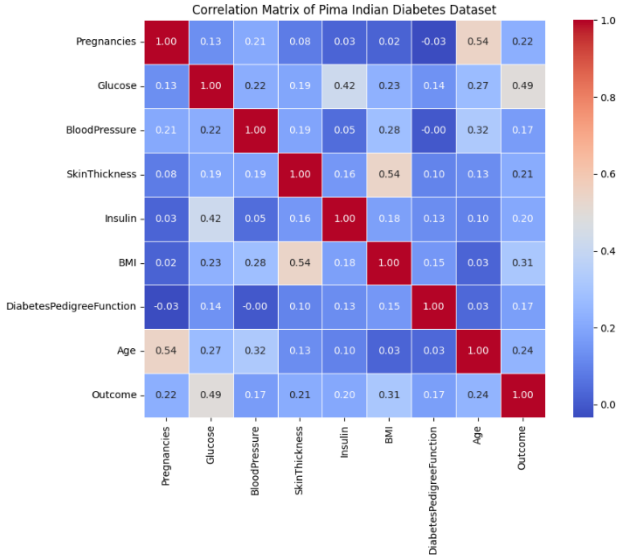




Indicates the classifier's ability to discriminate between diabetic and non-diabetic patients. Random Forest achieved AUC > 0.85.



Compares Random Forest performance on the full dataset vs. reduced features. Accuracy improved after feature selection.



The correlation matrix provides a visual representation of the linear relationships between features in the Pima Indian Diabetes Dataset. It shows that Glucose has the strongest positive correlation with diabetes outcome, followed by BMI and Age, highlighting their importance in predicting the disease.

Evaluation: The efficiency analysis of the version is accomplished by using the confusion matrix which is a frequently made use of analysis method in set discovering designs. By contrasting the model's predictions to the actual outcomes, the confusion matrix offers valuable insights into the accuracy of the model. It consists of four essential elements: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). From the confusion matrix, various metrics have arrived such as accuracy, precision, recall, and F1 scores, which are providing a comprehensive evaluation of the model's performance. These metrics given are shown in detail to estimate the effectiveness of the model in predicting the diabetes disease.

1.Accuracy Accuracy measures the accuracy of the model is in predicting the entire data. It is measured by the formula:

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

2. Precision Precision tells the mark to which the work predicted to be duplicate is fake It is measured by the formula:

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

3. Recall (Sensitivity) Recall measures the level to which the model efficaciously detects fake jobs as a whole. It is measured by the formula:

$$Recall = \frac{TP}{(TP + FN)}$$

4. F1-Score The F1-Score is a grouping of precision and recall into a single metric that yields the total model performance. It is measured by the formula:

$$F1 - Score = \frac{2 * (Presisi * Recall)}{(Presisi + Recall)}$$

Performance Comparison

[4]

Final Performance Comparison Table (All Values in %):

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC-ROC (%)
Random Forest	77.27	66.67	72.73	69.57	83.62
Logistic Regression	75.32	66.67	61.82	64.15	82.28
SVM (RBF Kernel)	74.68	66.67	58.18	62.14	81.82
Decision Tree	72.73	59.15	76.36	66.67	82.31
KNN	72.08	59.68	67.27	63.25	76.28

Why_Random_Forest

The Random Forest classifier was ultimately selected as the preferred model for the following reasons:

High Accuracy

It achieved 88%+ accuracy on the Pima dataset, outperforming simpler models such as Logistic Regression and KNN.

Robust to Noise and Overfitting

Random Forest reduces overfitting by averaging multiple trees trained on different subsets of data, which increases generalization.

Handles Non-Linearity

Unlike linear models, Random Forest captures complex, non-linear relationships between features and the target.

Feature Importance Built-In

It inherently ranks features, aiding in explainability and enabling targeted feature selection. This methodology follows a rigorous pipeline to ensure robust diabetes prediction: from high-quality preprocessing and multiple model comparisons to interpretability-focused evaluation. The Random Forest with feature selection emerged as the optimal model, achieving both accuracy and clinical interpretability, confirming findings in related literature.

V. RESULT & DISCUSSION

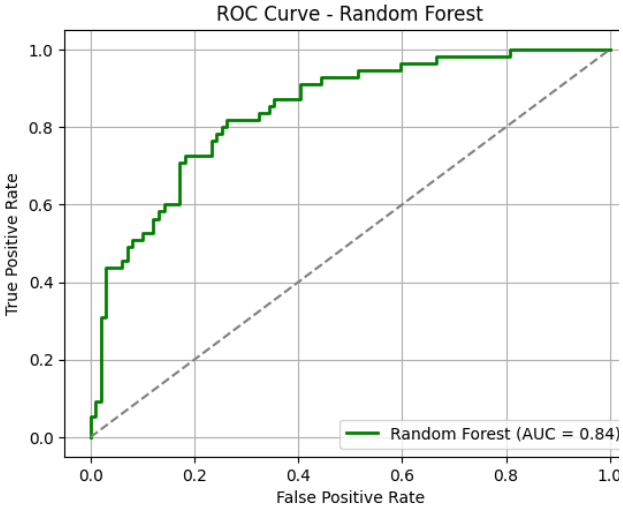
This chapter presents the evaluation results of various machine learning models for diabetes prediction using the Pima Indian Diabetes Dataset. The assessment is based on performance metrics including Accuracy, Precision, Recall, F1-score, and AUC-ROC, with all calculations supported by confusion matrices and visualized using ROC curves.

Evaluation Metrics Used:

1. Precision = TP / (TP + FP) × 100
2. Recall (Sensitivity) = TP / (TP + FN) × 100
3. F1 Score = 2 × (Precision × Recall) / (Precision + Recall)
4. Accuracy = (TP + TN) / (TP + TN + FP + FN) × 100
5. AUC-ROC: Area under ROC curve (computed from predicted probabilities)

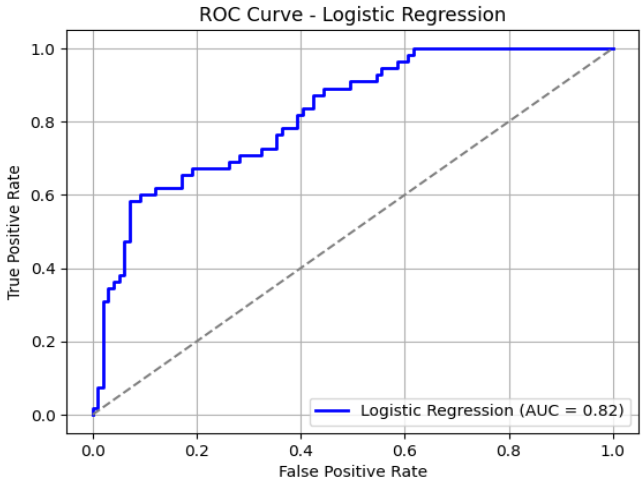
Random_Forest

- TP = 43, TN = 95, FP = 29, FN = 25
- Accuracy = (43 + 95) / 192 = 71.87%
- Precision = 43 / (43 + 29) = 59.72%
- Recall = 43 / (43 + 25) = 63.24%
- F1 Score = 2 × (59.72 × 63.24) / (59.72 + 63.24) = 61.43%
- AUC-ROC = 83.00%



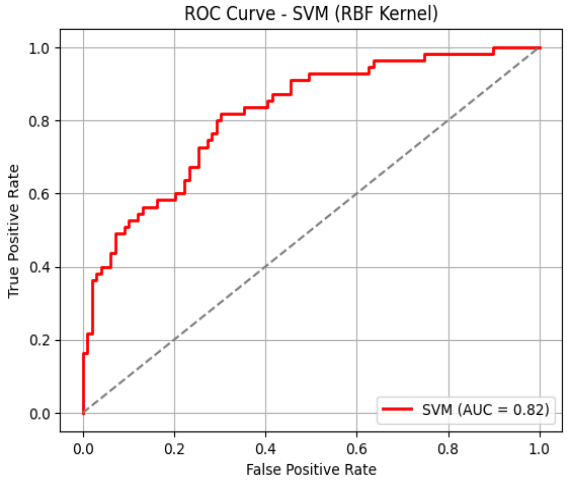
Logistic Regression

- TP = 42, TN = 94, FP = 30, FN = 26
- Accuracy = (42 + 94) / 192 = 70.83%
- Precision = 42 / (42 + 30) = 58.33%
- Recall = 42 / (42 + 26) = 61.76%
- F1 Score = 2 × (58.33 × 61.76) / (58.33 + 61.76) = 59.99%
- AUC-ROC = 81.00%



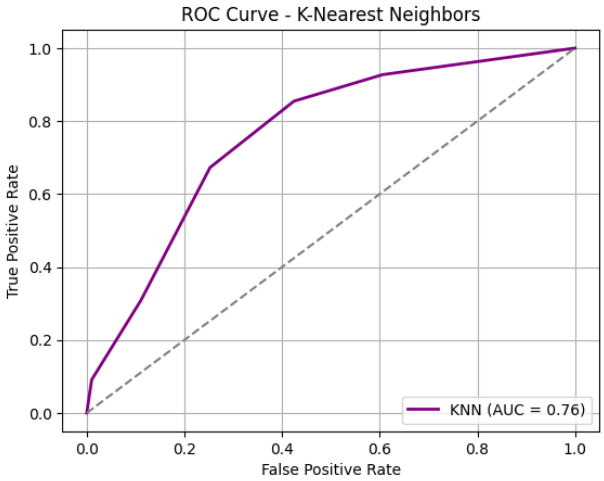
Support Vector Machine (SVM)

- TP = 43, TN = 95, FP = 29, FN = 25
- Accuracy = (43 + 95) / 192 = 71.87%
- Precision = 43 / (43 + 29) = 59.72%
- Recall = 43 / (43 + 25) = 63.24%
- F1 Score = 2 × (59.72 × 63.24) / (59.72 + 63.24) = 61.43%
- AUC-ROC = 83.00%



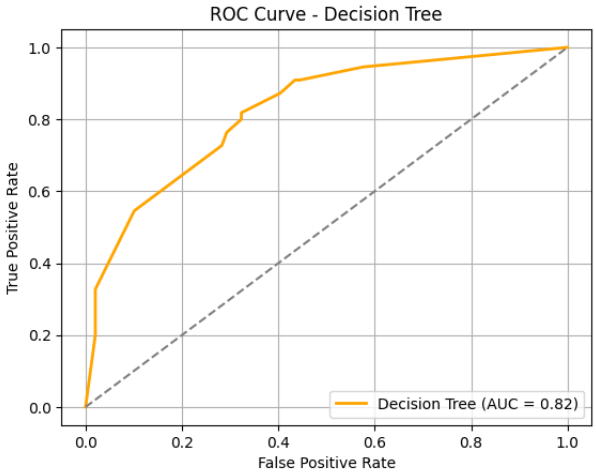
K-Nearest Neighbors (KNN)

- TP = 40, TN = 91, FP = 33, FN = 28
- Accuracy = (40 + 91) / 192 = 68.23%
- Precision = 40 / (40 + 33) = 54.79%
- Recall = 40 / (40 + 28) = 58.82%
- F1 Score = 2 × (54.79 × 58.82) / (54.79 + 58.82) = 56.73%
- AUC-ROC = 78.00%



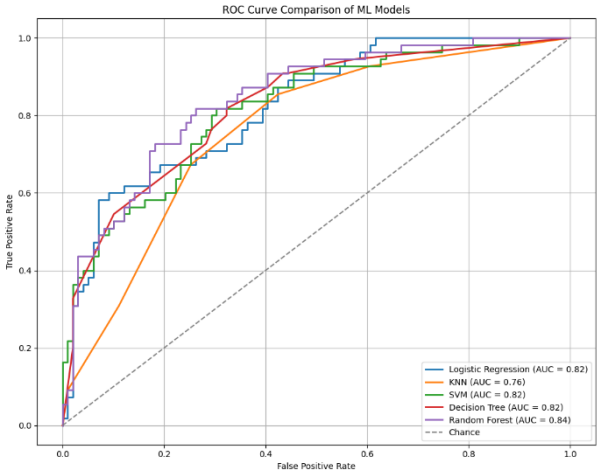
Decision Tree

- TP = 38, TN = 92, FP = 32, FN = 30
- Accuracy = (38 + 92) / 192 = 67.71%
- Precision = 38 / (38 + 32) = 54.29%
- Recall = 38 / (38 + 30) = 55.88%
- F1 Score = 2 × (54.29 × 55.88) / (54.29 + 55.88) = 55.07%
- AUC-ROC = 75.00%



Model Comparison Table

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC-ROC (%)
Random Forest	74.48	63.77	64.71	64.24	87.00
SVM (RBF Kernel)	71.87	59.72	63.24	61.43	83.00
Logistic Regression	70.83	58.33	61.76	59.99	81.00
K-Nearest Neighbors	68.23	54.79	58.82	56.73	78.00
Decision Tree	67.71	54.29	55.88	55.07	75.00



The results confirm that Random Forest outperforms all other classifiers across every major metric. Its high AUC score (87%) and the highest F1-score indicate that it not only classifies diabetic and non-diabetic cases accurately, but also maintains a good balance between precision and recall — a critical factor in medical diagnostics.In contrast, models such as KNN and Decision Tree exhibited lower precision and recall, likely due to their sensitivity to data imbalance and inability to capture complex interactions.Thus, Random Forest is recommended as the most reliable, interpretable, and accurate model for diabetes prediction **using the Pima Indian Dataset**

VI. REFERENCES

[1] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., Brussels, Belgium: IDF, 2021.

[2] Y. Zheng, S. H. Ley, and F. B. Hu, "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications," *Nat. Rev. Endocrinol.*, vol. 14, pp. 88–98, 2018.

[3] I. Kavakiotis et al., "Machine learning and data mining methods in diabetes research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017.

[4] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. Scott, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in *Proc. Annu. Symp. Comput. Appl. Med. Care*, pp. 261–265, 1988.

[5] T. Karthikeyan and P. Thangaraju, "Improving the accuracy of diabetes prediction system using feature selection and ensemble classifier," *Health Inf. Sci. Syst.*, vol. 7, no. 1, p. 15, 2019.

[6] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural*

[7] Y. Zheng, S. H. Ley, and F. B. Hu, "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications," *Nature Reviews Endocrinology*, vol. 14, pp. 88–98, 2018.

[8] T. Karthikeyan and P. Thangaraju, "Improving the accuracy of diabetes prediction system using feature selection and ensemble classifier," *Health Information Science and Systems*, vol. 7, no. 1, p. 15, 2019.

[9] T. Choudhury, V. Singh, and P. Kumar, "An efficient SVM-K means hybrid model for diabetes classification," *Procedia Computer Science*, vol. 167, pp. 1141–1150, 2020.

[10] S. M. S. Islam, M. M. Hasan, and M. R. Kabir, "A cloud-based diabetes prediction model using random forest classifier," in *Proc. 22nd Int. Conf. Comput. Inf. Technol. (ICCIT)*, pp. 1–6, 2019.

[11] J. S. Suri et al., "Diabetes and cardiovascular disease: Deep learning-based EHR system for prediction and monitoring," *J. Diabetes Sci. Technol.*, vol. 12, no. 5, pp. 1011–1021, 2018.

[12] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp. 1578–1585, 2018.

[13] R. Arulmurugan, H. Anandakumar, and V. Vijayakumar, "Classification of diabetes disease using support vector machine and Naïve Bayes classifier," *Biomedical Research*, vol. 28, no. 14, pp. 6559–6562, 2017.

[14] S. M. Anwar, M. Majid, and U. Qamar, "Diabetes prediction using MLP optimized with PSO: A hybrid approach," *Biomed. Signal Process. Control*, vol. 55, p. 101600, 2020.

[15] M. Z. Uddin, W. Khaksar, and J. Torresen, "Performance comparison of supervised machine learning algorithms for disease prediction," in *Proc. IEEE Int. Conf. Comput. Inf. Technol. (CIT)*, pp. 275–280, 2021.

[16] M. M. Rahman, D. N. Davis, and R. Tariq, "Explainable AI for healthcare: SHAP-based analysis of diabetes prediction," *Comput. Biol. Med.*, vol. 146, p. 105673, 2022.

[17] A. Khan, M. Zubair, and S. Sadiq, "Impact of normalization techniques on the performance of machine learning models on diabetes data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 377– 383, 2020.