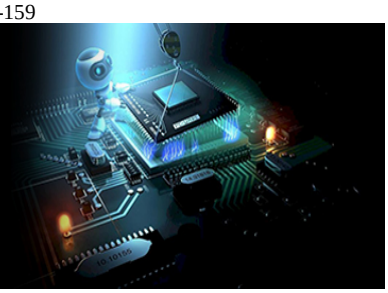
International Journal of Engineering in Computer Science



**E-ISSN:** 2663-3590
**P-ISSN:** 2663-3582
**Impact Factor (RJIF):** 5.52
www.computersciencejournals.com/ijecs
IJECS 2025; 7(2): 152-159
Received: 18-07-2025
Accepted: 20-08-2025

**Prashant Kaler**
Visvesvaraya Technological University / Department of CSE (MCA), Kalaburagi, Karnataka, India

**Dr. Swaroopa Shastri**
Visvesvaraya Technological University / Department of CSE (MCA), Kalaburagi, Karnataka, India

**Corresponding Author:**
**Prashant Kaler**
Visvesvaraya Technological University / Department of CSE (MCA), Kalaburagi, Karnataka, India

# Development of Accurate Machine Learning Models for Gestational Diabetes Prediction Using Patient Clinical Features

**Prashant Kaler and Swaroopa Shastri**

**DOI:** https://doi.org/10.33545/26633582.2025.v7.i2b.211

**Abstract**
Gestational Diabetes Mellitus (GDM) is a type of diabetes that develops during pregnancy and, if left untreated, can lead to serious complications for both mother and baby. Early detection is essential to reduce health risks, yet traditional screening methods such as the Oral Glucose Tolerance Test (OGTT) are usually performed late in pregnancy, limiting the time for preventive care. This study explores the use of machine learning (ML) techniques to predict the risk of GDM using clinical and demographic data. We used the well-known Pima Indians Diabetes Dataset, which includes eight key features: number of pregnancies, plasma glucose concentration, diastolic blood pressure, skin thickness, serum insulin, body mass index (BMI), diabetes pedigree function, and maternal age. Nine different ML algorithms—Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, AdaBoost, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, and XGBoost—were developed and tested. The dataset was divided into 70% training and 30% testing sets, and model performance was evaluated using Accuracy, Precision, Recall, and F1-score. Among these methods, the Random Forest model achieved the best overall results with an accuracy of 79% and a balanced F1-score of 0.66. Feature-importance analysis identified glucose level, insulin concentration, BMI, and number of pregnancies as the most significant predictors of GDM. These findings demonstrate that machine learning can be a valuable tool for early risk prediction, helping healthcare providers and expectant mothers take preventive steps before conventional screening is possible.

**Keywords:** Gestational Diabetes Mellitus, Machine Learning, Random Forest, Early Prediction, Clinical Features, Pima Indians Diabetes Dataset

## Introduction
Gestational diabetes mellitus (GDM) is a type of diabetes diagnosed during pregnancy that can lead to serious health complications for both mother and baby. Traditionally, GDM screening includes clinical tests like the oral glucose tolerance test (OGTT), usually performed between the 24th and 28th week of gestation. Although effective, this approach often diagnoses GDM relatively late, limiting the window for preventive interventions. Early prediction methods using conventional risk factors have often been limited by the complex interplay of various clinical and demographic factors [1]. In recent years, machine learning (ML) techniques have been applied to improve early detection by analyzing large and complex clinical datasets. ML algorithms can uncover subtle patterns and nonlinear relationships within features such as Plasma Glucose concentration, Blood Pressure, Skin Thickness, Insulin levels, Body Mass Index (BMI), Diabetes Pedigree Function, maternal Age, and Pregnancy history. For this study, these clinical features are taken from the well-known public Pima Indians Diabetes Dataset, which is widely used in academic research and tutorials for diabetes prediction modelling. These advances allow prediction of GDM risk earlier than conventional tests, aiding clinicians in proactive care and management [2, 3]. Importantly, the practical implementation of ML models offers significant benefits beyond clinical settings. Individuals with prior diabetes reports or risk assessments can input their relevant clinical data into an accessible predictive tool, receiving immediate feedback on their risk of developing gestational diabetes. This enables early awareness and lifestyle adjustment to reduce complications. Moreover, such technology reduces the reliance on costly and time-consuming laboratory tests for low-risk individuals, optimizing healthcare

resources and lowering overall diagnostic costs [4, 5]. This study aims to develop accurate machine learning models for GDM prediction using comprehensive patient clinical features. A variety of machine learning algorithms are applied to the dataset to identify which model performs best for early identification of gestational diabetes risk, thereby providing an effective and cost-efficient tool to support improved maternal and fetal health outcomes.

**Objectives**

To create predictive machine learning models for gestational diabetes mellitus (GDM) using clinical features including:

- Number of pregnancies
- Plasma glucose concentration
- Diastolic blood pressure
- Skin thickness
- Serum insulin levels
- Body mass index (BMI)
- Diabetes pedigree function (genetic predisposition)
- Maternal age
- To assess the performance of multiple machine learning algorithms and determine the most effective model for predicting GDM.
- To identify key clinical predictors that contributes most significantly to gestational diabetes risk.
- To provide a reliable tool for early screening of gestational diabetes that can assist clinicians in timely intervention and management.

**Literature review**

Diabetes is a major global health concern, affecting a significant proportion of the adult population and prompting extensive research into early detection and prediction. Over the years, a wide range of computational approaches have been explored to improve predictive accuracy, including machine learning (ML), neural networks (NNs), data mining techniques, and genetic algorithms. Among these, ML has emerged as a particularly powerful and versatile tool, gaining increasing attention in the medical research community. Its ability to process large datasets, analyze multiple risk factors simultaneously, and identify complex, non-linear relationships makes it well suited for clinical prediction tasks. Furthermore, advanced ML techniques provide effective feature-selection methods that can uncover hidden patterns and highlight the most influential variables. A growing body of literature demonstrates the effectiveness of ML models in predicting diabetes, consistently showing improved performance over traditional statistical methods. The following section reviews notable studies that have applied various ML algorithms to diabetes prediction, outlining their methodologies and key findings to provide a foundation for the present work.

In one notable study, [6] the researchers developed five predictive models based on different machine learning algorithms, including linear Support Vector Machine (SVM), multifactor dimensionality reduction, radial basis function, kernel SVM, K-Nearest Neighbors (KNN), and artificial neural networks. To ensure optimal feature selection, they employed the Boruta wrapper method, which effectively identifies the most significant variables in the dataset. Their experimental analysis revealed that all the models demonstrated promising performance in predicting diabetes. However, among the approaches tested, KNN and linear SVM achieved the highest accuracy in distinguishing diabetic from non-diabetic patients.

Similarly, Khanam and Foo [7], applied the Pima Indians Diabetes Dataset (PIDD) to evaluate the predictive capabilities of seven different machine learning models. As part of their feature selection strategy, they removed two less significant attributes to enhance model performance. Their results showed that Support Vector Machine (SVM) and Logistic Regression (LR) produced strong predictive accuracy for diabetes detection. In addition, they trained a Neural Network (NN) model on the same dataset using multiple hidden layers and varying the number of training epochs. The authors reported that, compared with the other approaches, a neural network configured with two hidden layers achieved the best overall performance.

A separate investigation carried out a meta-analysis of diabetes prediction techniques using machine learning (ML). In this study [8], the researchers applied the PROBAST (Prediction Model Risk of Bias Assessment Tool) framework to carefully evaluate the potential sources of bias within the ML-based models. To further examine heterogeneity and consolidate the results, the Meta-DisC software package was employed. The meta-analysis demonstrated that, overall, ML approaches provided superior predictive performance compared to conventional diabetes screening methods.

Another study focused on Logistic Regression (LR) as the primary predictive model while also exploring additional ML techniques such as Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) in various ensemble configurations to assess potential gains in performance. The experiments were conducted on two benchmark datasets: the Pima Indians Diabetes dataset, containing nine key attributes, and the Vanderbilt dataset, which includes sixteen features. The researchers further applied two different feature selection strategies to evaluate their impact on predictive outcomes. Their findings revealed that LR consistently produced strong and reliable results for diabetes prediction. Moreover, they highlighted that beyond the chosen algorithm, several other factors—such as feature selection methods and dataset characteristics—significantly influence the overall model accuracy [9].

In a related study [10], researchers explored multiple machine learning algorithms—including Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), AdaBoost (AB), XGBoost (XGB), and Naive Bayes (NB)—to enhance diabetes prediction using the Pima Indians Diabetes (PID) dataset. Their approach emphasized the importance of thorough preprocessing to ensure reliable and accurate outcomes. The experimental findings demonstrated that the best predictive performance was achieved by combining two boosting-based classifiers, AdaBoost and XGBoost, particularly when supported by an effective preprocessing pipeline. Similarly, another group of authors applied various ML techniques, such as Logistic Regression (LR), Decision Tree, Naive Bayes, Gradient Boosting (GB), Random Forest, KNN, and Support Vector Machine (SVM), to identify effective models for diabetes prediction. They highlighted the role of preprocessing steps—specifically label encoding and normalization—in improving model accuracy. Among the tested algorithms, SVM delivered the highest predictive performance. Additionally, the researchers employed several feature selection strategies to

identify and rank key risk factors, thereby enhancing both the interpretability and the predictive strength of their proposed models.

**Problem domain**

Gestational Diabetes Mellitus (GDM) is a type of diabetes that appears during pregnancy and can cause serious health problems for both mother and baby, such as high birth weight, preeclampsia, and a higher risk of developing Type 2 diabetes later in life. Traditional screening methods, like the oral glucose tolerance test (OGTT), are usually done between the 24th and 28th week of pregnancy. Because this is relatively late, it leaves little time for early lifestyle or medical interventions. These tests can also be expensive and less accessible in low-resource areas. GDM risk depends on many factors, such as glucose levels, blood pressure, BMI, maternal age, and family history, which make early prediction difficult with standard statistical methods.

Machine learning (ML) can handle these multiple risk factors at once and detect patterns that help identify women at risk much earlier. By using clinical features such as Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age, ML models can:

- give early risk predictions,
- guide timely lifestyle and medical interventions, and
- Reduce the need for costly laboratory tests.

This project focuses on building accurate and cost-effective ML models for early GDM prediction to improve maternal and fetal health outcomes.

**Methodology**

Gestational Diabetes Mellitus (GDM), early detection of remains a critical challenge, as traditional clinical assessments often achieve only 60% to 80% accuracy due to the complex interplay of multiple physiological factors. Subtle patterns in patient data can easily elude even experienced clinicians, making timely diagnosis difficult. Machine learning offers a promising solution by analyzing multifactorial datasets efficiently and providing reliable decision support with minimal human error. In this study, clinical features such as pregnancies, glucose levels, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, and age were used to train multiple machine learning models, including Logistic Regression, Random Forest, and another models. Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, AdaBoost, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, and XGBoost. These models were selected to explore both simple and complex relationships within the data, with Logistic Regression providing interpretability, ensemble methods like Random Forest and Gradient Boosting offering robustness and handling feature interactions effectively, SVM excelling in high-dimensional spaces, and XGBoost known for its speed and predictive accuracy [11, 12].

**Table 1:** Configured Hyperparameters for Machine Learning Models

| Model | Hyperparameter | Values Used or Range |
|---|---|---|
| Logistic Regression | Penalty | L2 |
| | Regularization Strength (C) | 1.0 (default) |
| | Solver | lbfgs |
| Random Forest | Number of Trees | 100 |
| | Max Depth | 20 |
| | Min Samples Split | 2 |
| | Max Features | sqrt |
| Decision Tree | Max Depth | 15 |
| | Min Samples Split | 2 |
| | Criterion | Gini |
| Gradient Boosting | Learning Rate | 0.1 |
| | Number of Estimators | 200 |
| | Max Depth | 5 |
| AdaBoost | Number of Estimators | 100 |
| | Learning Rate | 0.5 |
| K-Nearest Neighbors | Number of Neighbors | 7 |
| | Weights | Distance |
| | Metric | Minkowski |
| Support Vector Machine | Kernel | RBF |
| | C | 1.0 |
| | Gamma | Scale |
| Naive Bayes | - | No tunable parameters |
| XGBoost | Learning Rate | 0.1 |
| | Max Depth | 6 |
| | Subsample | 0.8 |
| | Colsample Bytree | 0.8 |

The above table 1: summarizes the key hyperparameters for each machine learning model used in this study. Hyperparameters are settings configured before training that control the learning process and impact model performance. Optimal hyperparameter selection is essential to balance model complexity and generalization. Values listed were chosen based on empirical validation and prior research to ensure effective training of each algorithm for gestational diabetes prediction.

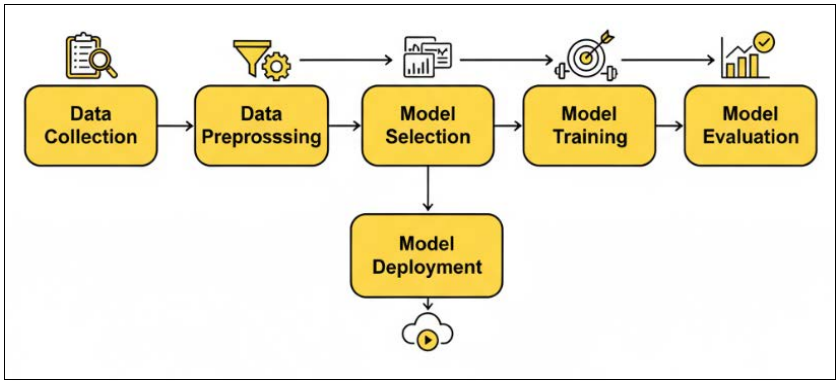



**Fig 1:** Proposed Methodology

In the above proposed methodology flow is represented with, data collection, where patient records were extracted from the publicly available Pima Indians Diabetes Dataset, widely used for academic research in diabetes prediction. The data consists of numeric clinical features relevant to GDM. Following collection**,** preliminary data cleaning and validation were performed to ensure quality and consistency. Missing values were handled using simple imputation techniques such as mean or median replacement, and inconsistencies or duplicate records were corrected. The dataset was then preprocessed using Python's pandas library for efficient data manipulation and preparation for machine learning training. For data visualization, Python libraries Matplotlib and Seaborn were used to generate informative plots and graphs, facilitating exploration of feature distributions, relationships, and patterns in the dataset. These visualizations supported a better understanding of the data and informed the subsequent model development [13, 14].

The selected machine learning models were trained and evaluated using the processed dataset, with stratified sampling and class balancing techniques applied to improve generalization and performance. Ensemble-based methods such as Random Forest and AdaBoost utilized bagging and boosting principles to reduce overfitting and enhance predictive accuracy. By training and comparing multiple algorithms, this study aims to identify the model that optimally balances accuracy, sensitivity, and interpretability, thereby supporting clinicians in the early and precise prediction of gestational diabetes. The approach highlights the advantages of leveraging diverse machine learning techniques to address complex healthcare prediction problems effectively [15].

Below table 2: shows the distribution of the 768 samples used in this study. Among these, 268 samples were diagnosed with diabetes while 500 were not diagnosed.

**Table 2:** Shows the distribution of the 786 samples used in this study

| Category | Number of Samples |
|---|---|
| Total Samples | 768 |
| Diabetes Diagnosed | 268 |
| No Diabetes Diagnosed | 500 |

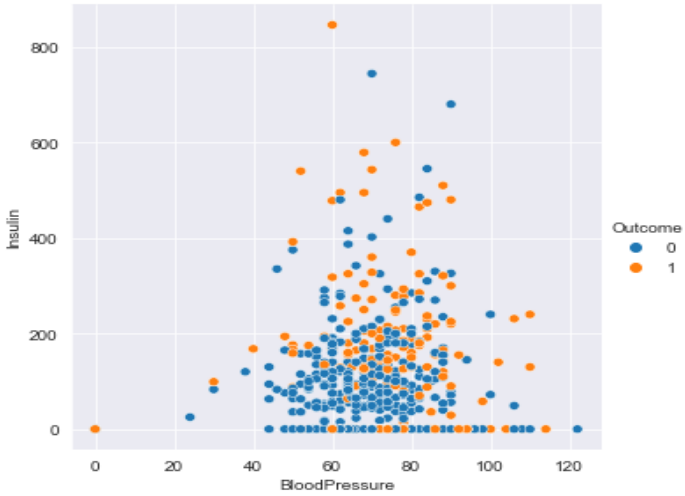
## Results and Discussion

This section presents the experimental findings of the proposed gestational-diabetes prediction framework and explains the evaluation of multiple machine-learning (ML) techniques. First, the clinical dataset was visually explored to highlight key patterns, after which the effectiveness of nine ML models was examined in detail. Finally, a comparative study of their predictive performances is reported together with a discussion of the hyper-parameters that were tuned for each model.

### Data-visualization insights

Initial exploratory analysis provided important clues regarding the relationship between the clinical features and the diabetes outcome. The age histogram demonstrated that the majority of subjects were between 20 and 40 years of age, coinciding with the age group at higher risk of gestational diabetes as shown in below Figure 2(a). The scatter plot of Blood Pressure versus Insulin revealed that patients diagnosed with diabetes (orange points) generally had higher insulin levels, whereas blood pressure values were widely spread and did not strongly distinguish the two classes as shown in Below Figure 2(b). A combined scatter plot of Glucose and Insulin values showed a clear positive relationship: women with higher glucose also tended to present with higher insulin levels, which is physiologically expected in diabetes-positive patients as shown in below Figure 2(c). The bar chart of diabetes prevalence by age group confirmed that women between 25 and 45 years were more frequently diagnosed, whereas cases beyond 45 years were rare as shown in below Figure 2(d). These visual findings emphasised the multifactorial nature of the problem and confirmed the relevance of using a variety of clinical indicators for model training.

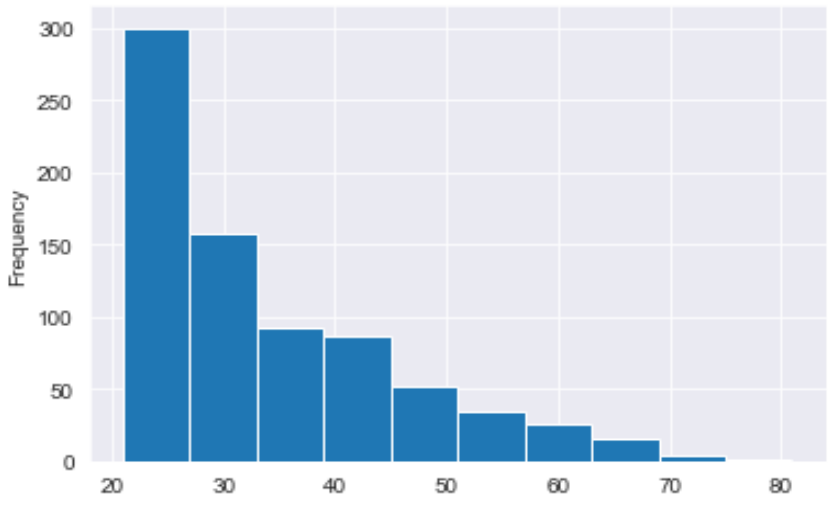
```
sns.relplot(x="BloodPressure", y="Insulin", data = df, hue="Outcome")
sns.set_style('darkgrid')
```




(a)

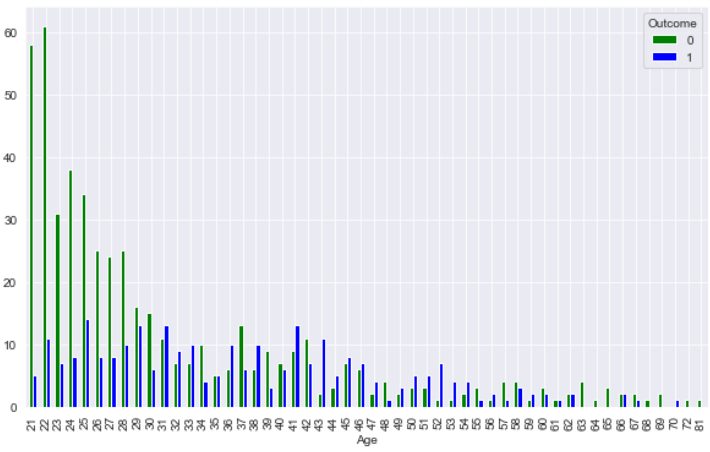
```
df.Age.plot.hist()
```

```
<AxesSubplot:ylabel='Frequency'>
```

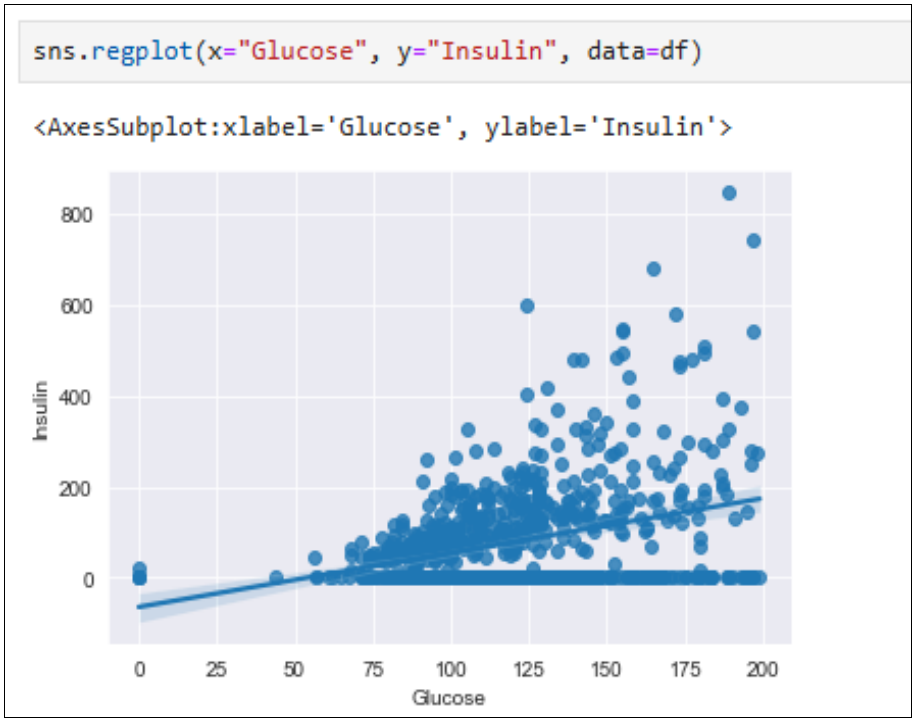



(b)

```
pd.crosstab(df.Age,df.Outcome).plot(kind="bar", figsize=(10,6), color=["green","blue"], width=0.6);
```




(c)




(d)

**Fig 2:** Exploratory data visualization of clinical parameters related to gestational diabetes diagnosis: (a) age distribution, (b) blood pressure vs insulin, (c) glucose vs insulin, and (d) diabetes prevalence by age group.

**Performance evaluation**

To measure classification quality, the dataset was split into 70% training and 30% testing sets using train_test_split(x, y, test_size=0.3, random_state=10) and the models were validated on the unseen 30% test data. Performance was quantified using the standard metrics Accuracy, Precision, Recall and F1-score, which are formally defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN},$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Here TP (True Positive) is the number of diabetic cases correctly predicted as diabetic, TN (True Negative) is the number of non-diabetic cases correctly predicted as non-diabetic, FP (False Positive) is the number of non-diabetic cases incorrectly predicted as diabetic, and FN (False Negative) is the number of diabetic cases incorrectly predicted as non-diabetic. Accuracy measures the overall proportion of correct predictions. These metrics provide complementary perspectives: Precision reflects the reliability of positive predictions, Recall measures how many actual positive cases are captured, and the F1-score balances the two.

**Hyper-parameter configuration of models**

To ensure a fair comparison, each algorithm was trained with carefully selected hyper-parameters:

- **Logistic Regression:** L2 regularisation with penalty='l2', solver=liblinear, C=1.0, maximum iterations = 100.
- **Random Forest:** 100 decision trees (n_estimators=100), maximum tree depth = None (allowing full growth), min_samples_split=2, bootstrap sampling enabled.
- **Decision Tree:** Gini impurity as splitting criterion, no restriction on maximum depth, minimum samples per leaf = 1.
- **Gradient Boosting:** n_estimators=100, learning rate = 0.1, maximum depth of each individual tree = 3, subsample = 1.0.
- **AdaBoost:** 50 weak learners (n_estimators=50) using decision stumps as base classifiers, learning rate = 1.0.
- **K-Nearest Neighbors (KNN):** n_neighbors=5, Euclidean distance metric, uniform weighting of neighbors.
- **Support Vector Machine (SVM):** Radial Basis Function (RBF) kernel with regularisation parameter C=1.0, kernel coefficient gamma='scale'.
- **Naive Bayes:** Gaussian Naive Bayes with default variance smoothing of 1e-9.
- **XGBoost:** n_estimators=100, learning rate = 0.1, maximum depth = 3, subsample = 0.8, column sampling by tree = 0.8.

These hyper-parameter settings were selected based on standard practices and preliminary tuning to balance bias and variance without overfitting.

**Comparative results of ML models**

**Table 3:** summaraises the classification performance of the nine models

| Classifier | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Logistic Regression | 0.77 | 0.67 | 0.62 | 0.64 |
| Random Forest | **0.79** | **0.74** | 0.59 | **0.66** |
| Decision Tree | 0.75 | 0.62 | 0.68 | 0.65 |
| Gradient Boosting | 0.75 | 0.64 | 0.60 | 0.62 |
| AdaBoost | 0.71 | 0.56 | 0.63 | 0.59 |
| K-Nearest Neighbors | 0.76 | 0.63 | 0.62 | 0.62 |
| Support Vector Machine | 0.76 | 0.63 | 0.62 | 0.62 |
| Naive Bayes | 0.75 | 0.63 | 0.62 | 0.62 |
| XGBoost | 0.77 | 0.72 | 0.54 | 0.62 |

The Random Forest classifier recorded the highest overall accuracy of 79% and the best F1-score of 0.66**,** indicating its ability to maintain a good balance between Precision and Recall.

Logistic Regression and XGBoost achieved competitive accuracies of 77%, though XGBoost's Recall was lower at 0.54, which slightly reduced its F1-score. Decision Tree produced a respectable Recall of 0.68—capturing many positive cases—but at the expense of slightly lower Precision. AdaBoost, with an accuracy of 71% and F1-score of 0.59, performed least effectively, suggesting that simple boosting of decision stumps was not sufficient for the complexity of this dataset.

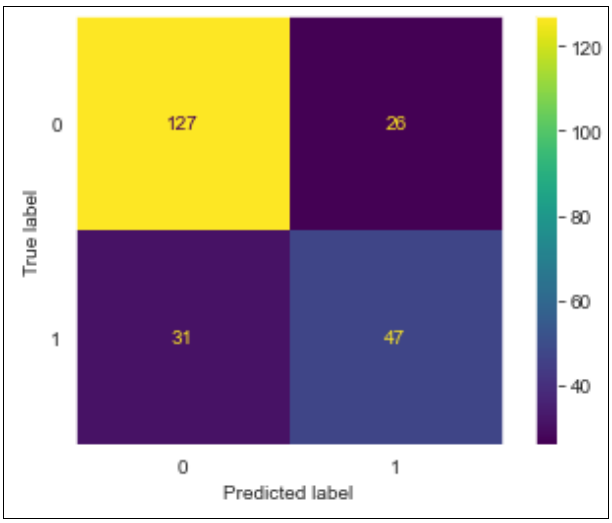



**Fig 3:** Confusion Matrix

The confusion matrix of the Random Forest model as shown in Figure 3 provides a detailed breakdown of its predictions, revealing high counts of both TP and TN with relatively few FP and FN. This confirms the model's ability to correctly classify both diabetic and non-diabetic individuals.

## Conclusion

Using the publicly available Pima Indians Diabetes Dataset, a range of supervised machine-learning algorithms were assessed for their ability to predict gestational diabetes risk. Among the tested models, Random Forest delivered the highest overall accuracy of roughly 79%, accompanied by balanced precision and recall values. The analysis also revealed that glucose concentration, insulin level, body-mass index, and pregnancy count were consistently the most influential variables, reinforcing established medical insights about key metabolic and obstetric risk factors. Although these outcomes demonstrate the promise of ensemble methods for early risk estimation, the results remain preliminary and drawn from a benchmark dataset. Future studies that incorporate real clinical records and more diverse populations will be essential to confirm the model's robustness and to prepare it for potential integration into healthcare decision-support systems.

## References

1. Mayo Clinic. Gestational diabetes - Symptoms and causes. 2025.
2. Kianian Bigdeli S, Ghazisaedi M, Ayyoubzadeh SM, Hantoushzadeh S, Ahmadi M. Predicting gestational diabetes mellitus in the first trimester using machine learning algorithms: a cross-sectional study at a hospital fertility health center in Iran. BMC Med Inform Decis Mak. 2025;25(3).
3. Watanabe M, Eguchi A, Sakurai K, Yamamoto M, Mori C, Japan Environment Children's Study (JECS) Group. Prediction of gestational diabetes mellitus using machine learning from birth cohort data of the Japan Environment and Children's Study. Sci Rep. 2023;13:17419.
4. Hu X. Prediction model for gestational diabetes mellitus using machine learning. Front Endocrinol. 2023;14.
5. Cleveland Clinic. Gestational Diabetes: Causes, Symptoms & Treatment. 2025.
6. Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. Appl Comput Inform. 2020;18(1–2):90–100.
7. Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express. 2021;7(4):432–9.
8. Zhang ZQ, Yang LQ, Han WT, *et al*. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J Med Internet Res. 2022;24(3):e26634.
9. Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. Comput Methods Programs Biomed Update. 2021;1:100032.
10. Hasan MK, Alam MA, Das D, Hossain E, Hasan M. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access. 2020;8:76516–31.
11. El-Sofany H, El-Seoud SA, Karam OH, Abd El-Latif YM, Taj-Eddin IATF. A proposed technique using machine learning for the prediction of diabetes disease through a mobile app. Int J Intell Syst. 2024;2024:6688934.
12. Kodama S, Fujihara K, Horikawa C, *et al*. Predictive ability of current machine learning algorithms for type 2 diabetes mellitus: a meta-analysis. J Diabetes Investig. 2022;13(5):900–8.
13. International Diabetes Federation. Diabetes Atlas. 2017.

Available from: http://www.diabetesatlas.org
14. Olisah CC, Smith L, Smith M. Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. Comput Methods Programs Biomed. 2022;220(12).
15. Aurélien G. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol (CA): O'Reilly Media, Inc.; 2021.
16. Mitchell TM. Machine Learning. New York (NY): McGraw-Hill; 2021.
17. Chatrati SP, Hossain G, Goyal A. Smart home health monitoring system for predicting type 2 diabetes and hypertension. J King Saud Univ. 2020;34(3):862–70.
18. Hasan MK, Alam MA, Das D, Hossain E, Hasan M. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access. 2020;8:76516–31.
19. Cervantes J, García-Lamont F, Rodríguez L, Lopez-Chau A. A comprehensive survey on support vector machine classification: applications, challenges, and trends. Neurocomputing. 2020;408:189–221.