Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 09 Issue 01
2025

*Research Article*
https://doi.org/10.56979/901/2025

# Prediction of Diabetes Disease during Pregnancy Using Machine Learning

**Rehana Yousaf[1], Uswah Farooq[2], Erssa Arif[3*], Smavia Neelam[4], Muhammad Junaid Ahsan[5], and Muhammad Amjad[1]**

[1]Department of Computer Science, Riphah International University, Faisalabad, Pakistan.
*Corresponding Author: Dr. Erssa Arif. Email: dr.erssa@riphahfsd.edu.pk

**Received:** April 25, 2025 **Accepted:** May 29, 2025

**Abstract:** In recent years, diseases associated with diabetes have emerged as one of the leading causes of death globally. In today's world, diabetes accounts for approximately one death every minute. This research mainly aims to determine which patients are at higher risk of developing diabetes based on different medical traits. A significant goal of this study is to create a system for disease awareness and prediction. Recently, various researchers have employed machine learning methods to aid healthcare professionals and specialists in diagnosing diabetes. This research highlights the use of supervised learning methods, such as Linear Regression, Naive Bayes, Decision Trees, and Logistic Regression, with the proposed model reaching an accuracy rate of 86%. The research concentrates on five significant parameters associated with diabetes: blood sugar level, blood pressure, cholesterol, uric acid, and hypertension. These factors are evaluated to understand their influence on the onset of diabetes. The findings illustrate successful outcomes using these algorithms. While diabetes is not curable, it can be effectively managed if detected early. Machine Learning (ML) facilitates the automated early identification of diabetes and has shown to be more efficient and precise compared to conventional human diagnosis.

**Keywords:** Clinical Decision Support; Diabetes Prediction; Early Diagnosis Systems; Feature Selection Techniques; Machine Learning Models; Maternal Health Analytics; Risk Factor Analysis

## 1. Introduction

Diabetes mellitus, often simply called diabetes, is a metabolic condition marked by the improper management of blood glucose (BG) levels, resulting in both immediate and long-term health issues. Its consistently rising incidence is impacting a growing number of families globally. There are primarily two varieties of diabetes. Type 1 diabetes generally appears in children and is triggered by an autoimmune reaction that hinders the pancreas from generating insulin, leading to a total absence of insulin. Type 2 diabetes, which is more prevalent among adults and the elderly, is defined by insulin resistance—where the body either does not produce sufficient insulin or cannot utilize it effectively.

If diabetes is not diagnosed or managed properly, it can heighten the likelihood of various serious health issues, including cardiovascular disease (CVD), nephropathy, retinopathy, and neuropathy. It also plays a significant role in causing microvascular damage [1]. In recent years, advancements in healthcare have been made in disease prevention, especially in alleviating the impact of CVD through enhanced diagnostic methods. CVD continues to be a major cause of mortality, affecting both men and women nearly equally. However, during this current decade, women seem to be more affected, with increasing evidence supporting this observation. Women often encounter additional health difficulties such as diabetes during pregnancy and gestational hypertension, both of which are associated with CVD.

Women are more likely to have other risk factors, such as migraines, which are also linked to an increased risk of heart attacks, albeit these are not unique to women. The American Stroke Association (ASA) claims that these risk factors unique to women have prompted the development of more specialized methods for assessing the risk of CVD [2]. With 37% of men and 38% of women under 75 dying from CVD in Europe, the disease's pervasiveness is evident.

According to a recent survey, Pakistan has improved life expectancy over the last three decades, going from 61.1 to 65.9 years. Reductions in infectious, maternal, neonatal, and nutritional disorders are responsible for this improvement. Even though life expectancy has increased by 8% for women and 7% for men over the past 30 years, Pakistan's life expectancy is still 7.6 years lower than the global average. The research also underlined that Pakistan has improved population health overall and is still looking for creative answers to healthcare and policy issues, even though the country has experienced political and economic upheaval since 1990.The primary objectives of this study are:

1. To develop a highly accurate model for heart disease prediction.
2. To identify pre-diabetic individuals using machine learning algorithms and key health parameters.
3. To enhance a data-driven decision support platform for more effective diabetes management.

## 2. Literature Review

A targeted research of hospitalized mood disorder patients from April to December 2015 was carried out. Their research revealed a crucial connection between a higher risks of heart failure, which impairs the heart's capacity to pump blood effectively, and decreased cardiac wall strength. Breathing becomes more challenging due to the inadequate blood supply caused by this illness. Numerous conditions, such as coronary artery disease, diabetes, high plasma volume, substance abuse (such as cocaine or alcohol), HIV, thyroid disorders, vitamin E insufficiency, and chemotherapy or radiation exposure, can cause heart failure [3].

### 2.1. Methods

Patients over 40 with a diagnosis of left ventricular systolic dysfunction who were all assigned to NYHA Class III or IV were included in the study. Using factors such age, hospital discharge status, serum creatinine, sodium levels, anemia, platelet count, creatinine phosphokinase, blood pressure, gender, diabetes, and smoking status, Cox regression analysis was utilized to model the risk of death. An early high death rate was followed by a gradual rise over time, according to Kaplan-Meier survival analysis. The functional form of covariates was evaluated using martingale residuals. A monogram was developed for visual survival probability prediction, and calibration slope analysis was used to confirm the model's discrimination and calibration. Anemia, blood pressure, ejection fraction, age, and renal failure were found to be important risk factors for mortality [4].

### 2.2. Predictive Modeling for Hospitalization Risk

Based on these results, the authors created a prediction model to determine which patients will require re-hospitalization within a year as a result of Type II diabetes problems. In order to explain hospitalization risks, cluster analysis was utilized to combine patients with comparable features, including demographic traits like race and ethnicity. These clusters were frequently discovered in diagnostic data's lower-dimensional subspaces. Since a tiny fraction of patient data was found to be highly predictive of future hospitalizations, electronic health records, or EHRs, were very helpful. In order to create sparse linear SVM classifiers that could reliably distinguish between hospitalized and non-hospitalized individuals, it investigated a variety of supervised machine learning classification approaches. During the hospital sessions, their approach also identified patient groupings that had not been observed before [5].

### 2.3. IoT and Cloud-Based Cardiac Prediction

Cardiac problems were the focus of the recently announced cloud- and IoT-based disease prediction algorithm. They developed an effective heart illness prediction system using sensor data and the UCI heart disease dataset. This dataset was used to train classification methods, which allowed the model to anticipate cardiac problems with fresh patient data. In terms of accuracy and applicability for IoT-enabled healthcare models, the J48 method fared better than the other classifiers examined, including MLP, SVM, LR, and J48 [6].

### 2.4. Classification Algorithm Evaluation

NB, KNN, LR, DT, NN, SVM, and a voting ensemble approach were the seven classification algorithms that were tested. Using 10-fold cross-validation, the Cleveland dataset—which has 303 records and 76 attributes—from the UC Irvine Machine Learning Repository was utilized for both training and testing.

Unbalanced training-to-testing ratios were less of a problem because to this strategy. With an accuracy of 87.5%, the Voting Classifier was the most accurate [7].

### 2.5. Cardiovascular Disease Prediction and Model Performance

Cardiovascular disease (CVD) is now the world's leading cause of death, it was underlined. The two most important methods for lowering mortality are still early diagnosis and efficient intervention. Healthcare professionals can use extensive patient health data to inform their judgments by utilizing Data Mining (DM) and Machine Learning (ML).

Their research, which centered on the Cleveland dataset, emphasized the application of supervised learning methods like decision trees and naive bayes. Only 14 of the 76 variables in the sample were determined to be essential for predicting heart disease with any degree of accuracy. The K-Nearest Neighbor (KNN) approach demonstrated the highest prediction accuracy out of all the algorithms that were tested. Subsequent research trained a model on the same Cleveland dataset using a multilayer perceptron neural network (MLP-NN) with back propagation. Performance was evaluated using key evaluation criteria, including recall, accuracy, precision, and F-measure. Following the selection of the 14 most pertinent features and preprocessing to deal with missing values, MLP-NN ran in just 3.86 seconds and obtained an accuracy of 93.39% [8].

## 3. Proposed Methodology

The methodology uses in this research Random Forest (RF), SVM, KNN, Logistic Regression, Naive Bayes and Decision algorithm to predict the cardiac diseases.

### 3.1. Proposed System

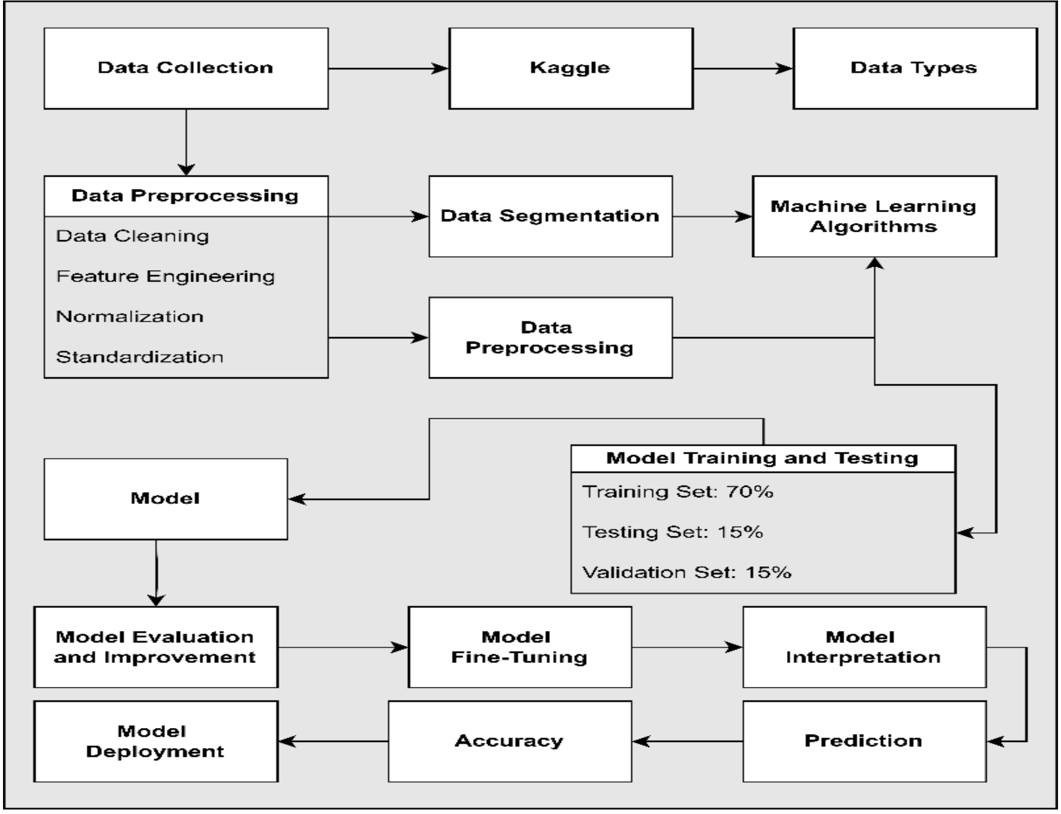

**Figure 1.** Overall work flow of purposed methodology

### 3.2. Supervised Learning Algorithms

### *3.2.1. K-Nearest Neighbor (KNN)*

The K-Nearest Neighbor (KNN) algorithm is a simple yet powerful supervised machine learning technique. It classifies data points based on the similarity to their nearest neighbors within the training dataset. Using Euclidean distance, the algorithm identifies the *K* closest data points to a given input and predicts the output based on the majority class among these neighbors.

Although effective, KNN can be computationally intensive as it requires searching through the entire dataset to make predictions, especially when working with large datasets or multiple features. To improve performance, it is beneficial to preprocess the data by scaling features and selecting an appropriate value of K and for the distance metric. The techniques such as cross-validation and the feature selection are also commonly used to optimize model's accuracy and efficiency.

### *3.2.2. Decision Tree*

A decision tree is a structure that resembles a flowchart, with internal nodes standing in for decision points based on input attributes, branches for the decisions' results, and leaf nodes for the final classification or prediction. At each stage, the algorithm recursively divides the dataset into subgroups according to the most instructive feature.

In order to identify which feature has the best predictive value and produces the least amount of impurity, the splitting criterion frequently uses measures like the Gini Index or Information Gain. Decision trees can handle both numerical and categorical data and are simple to understand and apply. They are reasonably resistant to outliers and efficient in identifying non-linear correlations. However, if improperly pruned, they could be prone to overfitting.

### *3.2.3. Naive Bayes*

Naive Bayes is a probabilistic classification technique based on Bayes' Theorem, which assumes independence among predictors. Despite the simplicity of this assumption—often unrealistic in real-world scenarios—the model performs well in many applications, especially in text classification and spam filtering.

The algorithm calculates the probability of a data point belonging to a particular class based on the conditional probabilities of its features. For categorical data, these probabilities are computed by dividing the count of feature instances by the count of instances within each class. Naive Bayes is known for its fast training and prediction times, making it ideal for large-scale problems.

### *3.2.4. Support Vector Machine (SVM)*

With their roots in statistical learning theory, Support Vector Machines (SVMs) are sophisticated supervised learning algorithms. They frequently beat conventional algorithms in terms of accuracy and are especially useful for classification issues. SVMs solve a constrained quadratic optimization problem in order to determine the appropriate hyper plane in the feature space for class separation. SVMs use kernel functions (such as polynomial and radial basis functions) to convert the input space into a higher-dimensional space where a linear separator can be discovered when the data is not linearly separable. Unlike neural networks, SVMs optimize a convex cost function, reducing the risk of getting stuck in local minima. This property, along with their foundation in Structural Risk Minimization (SRM), helps SVMs avoid overfitting, which is common in models based on Empirical Risk Minimization (ERM). However, a notable limitation of SVMs is that they provide strictly binary classification results and do not offer probability estimates or support for multi-class classification without additional methods.

### *3.2.5. LR (Logistic Regression)*

One popular supervised learning approach that works well for binary classification applications is logistic regression. Using one or more input features, it forecasts the likelihood of a binary result. The sigmoid function, sometimes referred to as the logistic function, and is the hypothesis function used in logistic regression as opposed to linear regression. This function is appropriate for modeling binary outcomes since it converts real-valued inputs to a range between 0 and 1. The method of logistic regression is straightforward, efficient, and simple to understand. It manages feature-target variable relationships that are both linear and somewhat non-linear. However, unless regularization or feature engineering are used to extend it, its performance can deteriorate when complex, high-dimensional interactions are present.

## 3.3. Data Preprocessing

Data preprocessing refers to the process of preparing raw data for further analysis and model training by eliminating incomplete, inconsistent, or unstructured data. This step is essential in any machine learning (ML) workflow and typically marks the beginning of the process. After analyzing the data, valuable insights can be drawn that help inform the development of ML models. Effective data preprocessing ensures that machine learning and statistical models can be developed more efficiently and yield reliable results. It includes various operations such as data cleaning, normalization, transformation, and feature selection. These steps are crucial for ensuring accurate, trustworthy, and robust enterprise-level applications in data science and artificial intelligence. Real-world data is often unorganized and may contain missing values, manual entry errors, duplicates, or inconsistent naming conventions. While humans can sometimes identify and correct these issues manually, ML algorithms require that the data be preprocessed automatically to function effectively. The data may include both continuous and categorical variables, requiring transformation to a uniform format suitable for algorithmic interpretation. Preprocessing involves steps such as removing blank fields, normalizing values, and eliminating outliers. Since ML algorithms require data in a specific format, preprocessing converts raw inputs into a format machines can understand. The goal of this stage is to transform unprocessed data, which may exist in formats such as text files, Excel sheets, or other structured files, into refined datasets with relevant and usable features.

### 3.4. Data Cleaning

Raw data often contains noise—random errors or discrepancies in values—that must be addressed before model training. Data cleaning involves handling missing values, identifying and correcting inconsistencies, and removing or minimizing noise and outliers. Unclean data leads to poor model performance and unreliable predictions. Thus, data cleaning is a vital step in preventing inaccurate patterns or misleading results during analysis. This process ensures that the dataset is suitable for mining and modeling tasks.

### 3.5. Data Integration

Data integration is the process of combining data from multiple sources into a single, unified dataset, typically stored in a data warehouse or data cube. When done carefully, it reduces redundancy and inconsistency, ensuring that subsequent data mining and analysis tasks are faster and more accurate.

### 3.6. Data Reduction

Raw datasets often contain redundant or irrelevant information, making them inefficient for ML and analytics tasks. Data reduction techniques, such as Principal Component Analysis (PCA) and feature selection, simplify the dataset by removing unnecessary attributes. This helps improve algorithm efficiency and clarity by focusing only on the most relevant features. Data reduction can include aggregating records, filtering unnecessary columns, and transforming the dataset into a compact yet informative form suitable for modeling.

### 3.7. Data Loss

During data preprocessing, efforts are made to minimize information loss while reducing data dimensionality. Techniques such as feature selection and dimensionality reduction aim to retain as much valuable data as possible, ensuring that critical patterns and trends remain intact for accurate modeling.

### 3.8. Training Machine Learning Algorithms

After preprocessing, the cleaned and structured data is fed into several machine learning algorithms for training. In this study, five primary algorithms were used: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), and Random Forest (RF). These methods were employed to analyze the dataset and generate classification predictions. This stage also helps uncover hidden patterns and insights related to the presence of heart disease. The training process uses different combinations of training and testing datasets (e.g., 80% training and 20% testing). These algorithms were assessed for accuracy and performance after preprocessing. Notably, Naive Bayes (NB) has proven effective for classification tasks in multiple studies (Al Reshan et al., 2024).

### 3.9. Model

Once the training phase is complete, the machine learning model is ready for testing and deployment.

#### 3.9.1. *Disease Prediction*

The trained model is evaluated by comparing input symptoms with known patterns to predict the likelihood of heart disease. At this point, the model estimates whether or not a patient is at risk.

#### 3.9.2. *Accuracy Measurement*

The performance of the trained model is validated using accuracy metrics. If the model meets or exceeds the required accuracy threshold, it is considered suitable for real-world application. All results, evaluation reports, and operational details of the proposed model are discussed in this chapter. Python-based tools were used to implement and assess the accuracy of each algorithm. First, the individual results of each algorithm are presented, followed by a comparative analysis. For clarity, data is summarized in tables and visualized through graphs to provide a quick and intuitive overview of the outcomes.

## 4. Dataset analysis

### 4.1. Bar Chart

Bar plots, also known as bar charts, are used to display categorical data using rectangular bars. The height or length of each bar represents the value or frequency of a specific category. All bars are drawn from a common baseline to allow easy comparison between categories. Bar charts are particularly effective for visualizing categorical variables, which take on discrete values that function as labels or distinct groups.

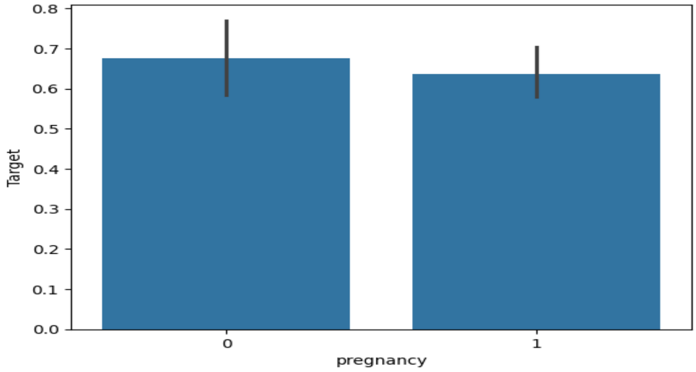

**Figure 2.** Bar Plot of Pregnancy

The "Pregnancy" variable serves as a proxy to indicate the patient's gender, with 0 representing females and 1 representing males. A bar chart in Figure 2 depicts the gender-based distribution of patients. The relative proportions of male and female patients in the dataset can be determined by comparing the bar heights. This visualization could indicate greater susceptibility to heart disease in either females or males within the specified population.

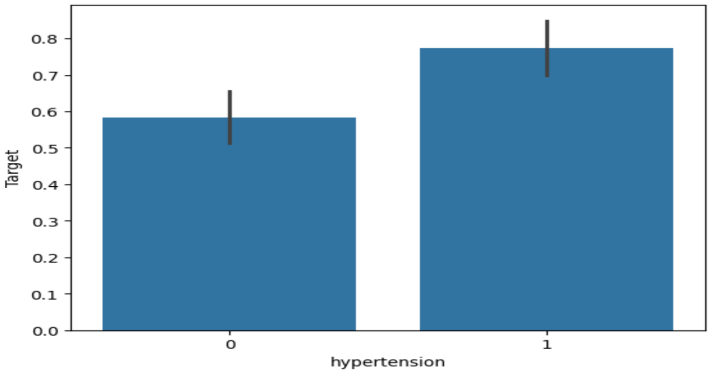

**Figure 3.** Distribution of the "Hypertension" Variable

*Figure* 3 illustrates the distribution of patients based on the "hypertension" variable, where 0 indicates the absence of hypertension and 1 indicates its presence. This bar chart allows for the assessment of hypertension prevalence within the patient population. By analyzing the relative heights of the bars for each category, one can evaluate how common hypertension is among the individuals in the dataset (Yu et al., 2024).

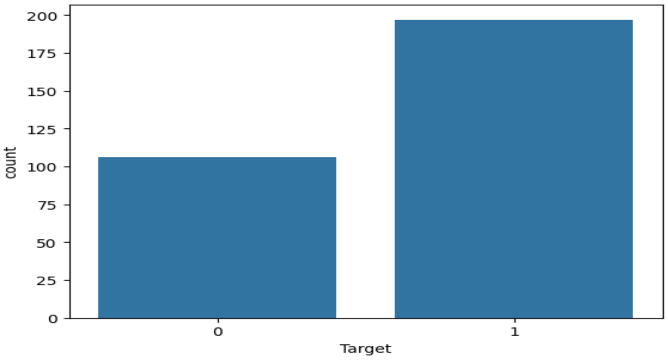

**Figure 4.** Distribution of the "Target" Variable

Figure 4 represents the "target" variable, where 0 indicates the absence of heart disease and 1 indicates its presence. The bar chart displays the distribution of patients diagnosed with and without heart disease. This visualization facilitates a comparison of the proportions in each category, offering insights into the overall prevalence of heart disease within the patient population.

### 4.2. Histogram

Histograms are visual tools designed to display the distribution of numerical data.

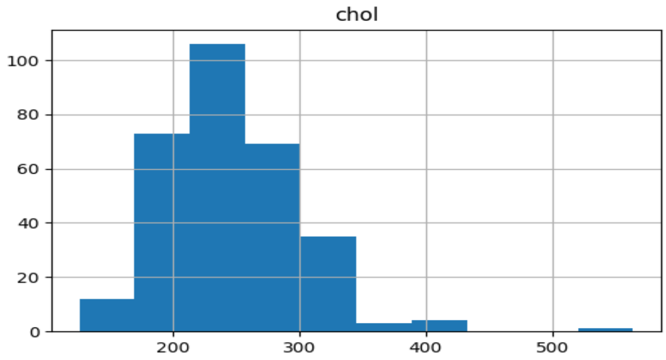

**Figure 5.** Histogram of Cholesterol (Chol)

Figure 5 displays a histogram showing the distribution of cholesterol levels across different intervals. Each bar represents a specific cholesterol range, with its height indicating the number of individuals within that range. By examining the intervals with the highest frequencies, one can identify the most common cholesterol levels in the dataset. Additionally, if higher cholesterol ranges show increased frequencies, this may suggest a potential correlation between elevated cholesterol levels and the presence of heart disease.

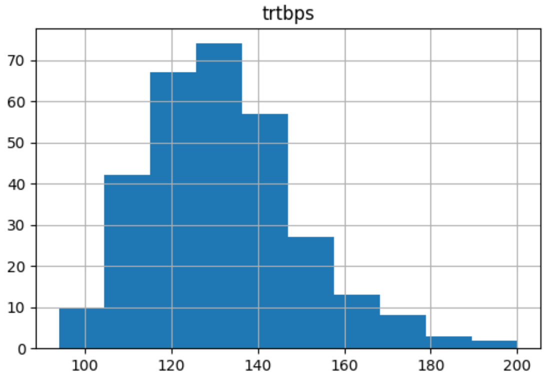

**Figure 6.** Histogram of Resting Blood Pressure (trtbps)

Figure 6 presents a histogram illustrating the distribution of resting blood pressure values across different intervals. Each bar reflects a specific range of blood pressure, with its height representing the number of individuals in that range. By identifying the intervals with the highest frequencies, one can determine the most common resting blood pressure levels among the patients. Additionally, the histogram may reveal patterns—such as higher frequencies at extreme blood pressure levels—that could indicate a potential relationship between abnormal resting blood pressure and heart disease (Gui et al., 2024).

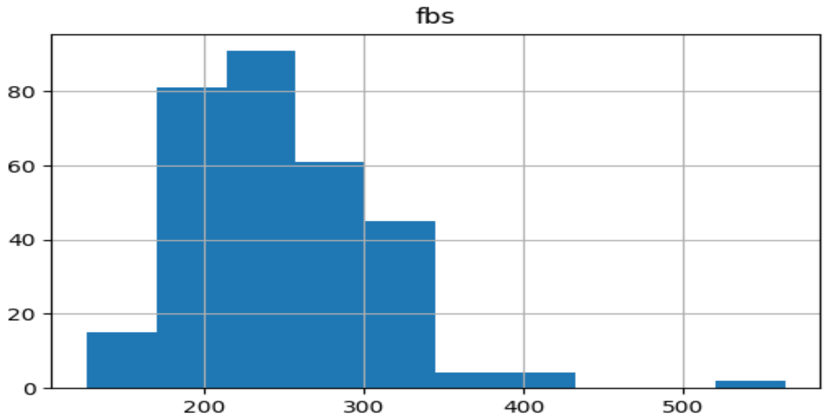

**Figure 7.** Histogram of Fasting Blood Sugar (fbs)

Figure 7 illustrates the distribution of fasting blood sugar levels through a histogram. Each bar represents a specific interval of blood sugar values, with the height indicating the number of individuals in that range. By analyzing the frequency of values across different intervals, patterns or variations in the distribution can be identified. The chart also allows for an assessment of whether differences in fasting blood sugar levels correspond to varying rates of heart disease prevalence (Zhang et al., 2024).

### 4.3. Relationships between Numerical Variables

Figure 8 displays a pair plot that visualizes the relationships between all pairwise combinations of numerical variables in the dataset. These scatter plots enable a closer examination of potential correlations and trends among features. For example, the relationship between cholesterol levels and other variables—such as resting blood pressure or fasting blood sugar—can be evaluated to detect any associations, such as higher cholesterol levels correlating with elevated blood pressure or glucose levels. Similarly, patterns involving fasting blood sugar and its relationship with cholesterol or blood pressure can be observed. The chart also facilitates analysis of resting blood pressure in relation to other variables, helping identify any trends that suggest higher blood pressure levels might align with increased cholesterol or fasting blood sugar levels.

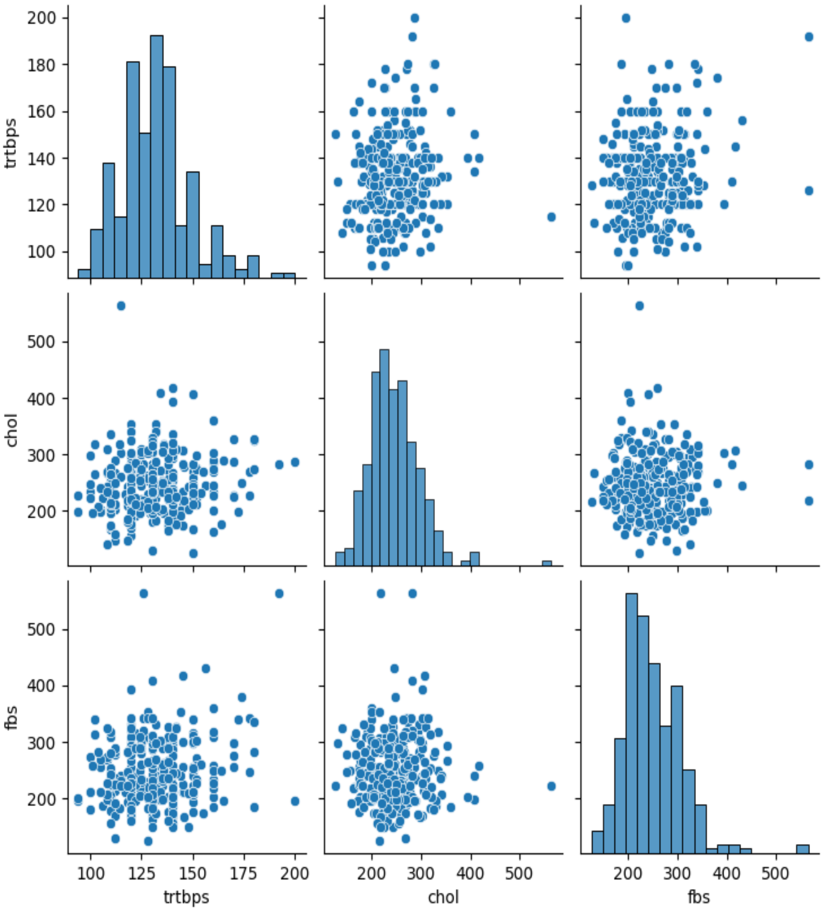

**Figure 8.** Pair Plot of Numerical Dataset

### 4.4. Representing the Presence or Absence of Heart Disease

The help reduce mortality rates caused by cardiac conditions, recent research has emphasized the importance of early detection of heart disease. A novel predictive model has been developed to analyze the relationships between cardiovascular risk factors and other variables. This model aims to improve the accuracy of heart disease prediction. By utilizing these findings, earlier identification of cardiovascular issues becomes possible, enabling timely intervention and potentially lowering heart disease-related death rates. This study also contributes valuable insights into how various cardiovascular factors are associated with other health indicators.

Figure 9 presents a pair plot illustrating the distinction between individuals with and without heart disease within the dataset. The "Target" variable is used as the hue parameter, meaning data points are color-coded based on whether heart disease is present (1) or absent (0). This visualization helps reveal variations in data distribution, cluster formations, and distinct relationships between variables across the two groups. These observed patterns may highlight variables that serve as significant predictors of heart disease (Xing et al., 2024).

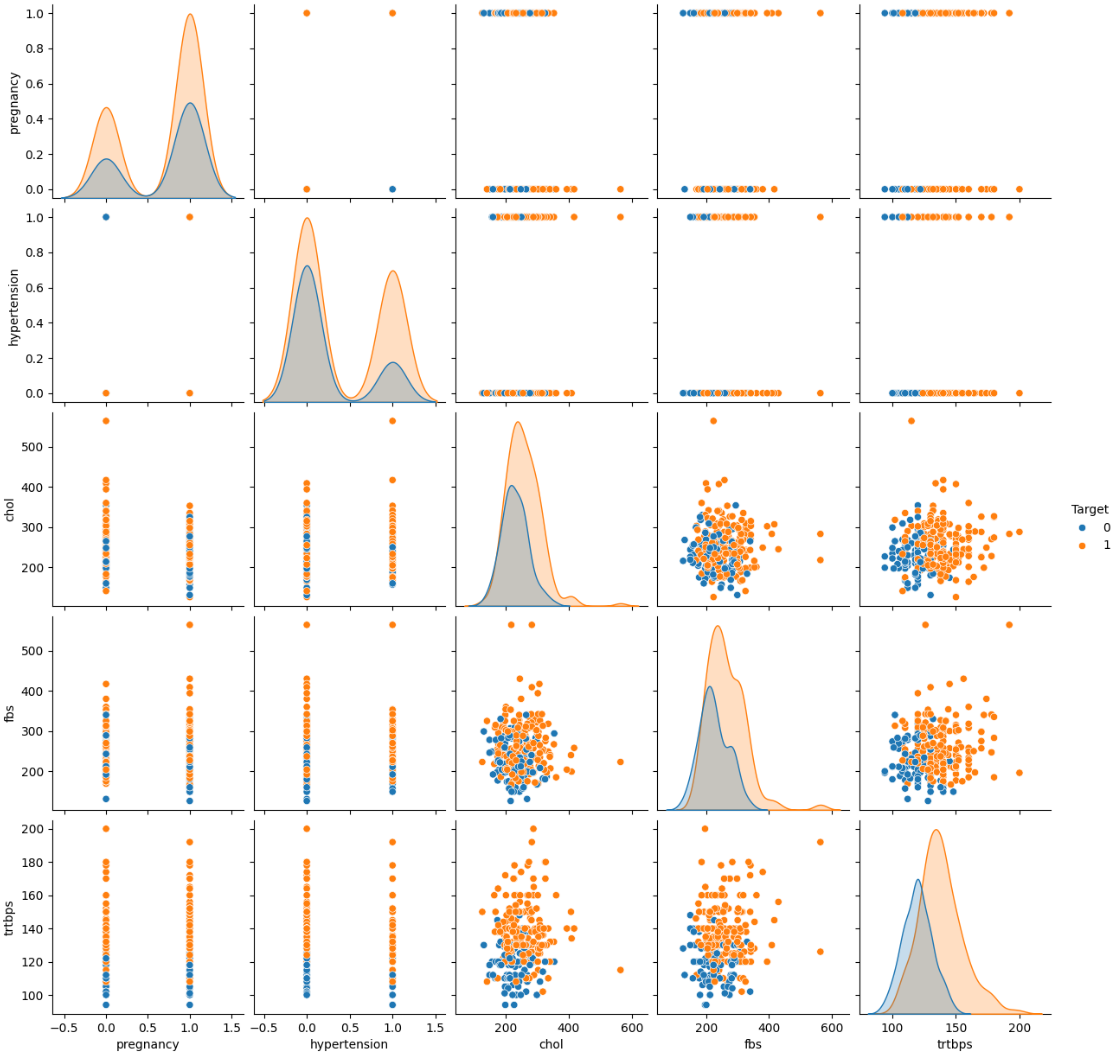

**Figure 9.** Visualization of Dataset With and Without Heart Disease

### 4.5. Correlation Heat map

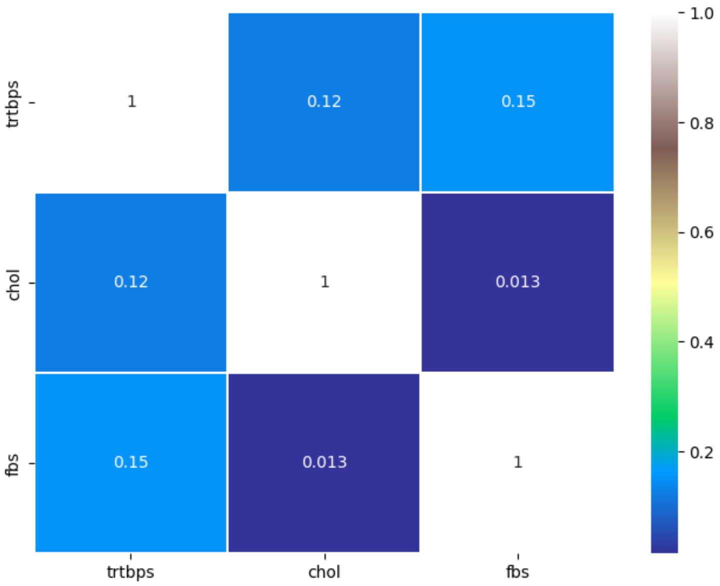

**Figure 10.** Heatmap of Correlation Matrix

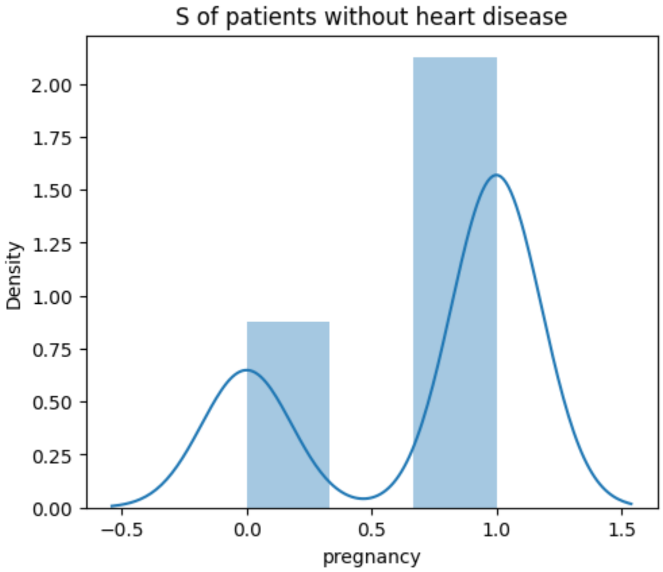

**Figure 11.** Pregnancy Patience without Heart Disease

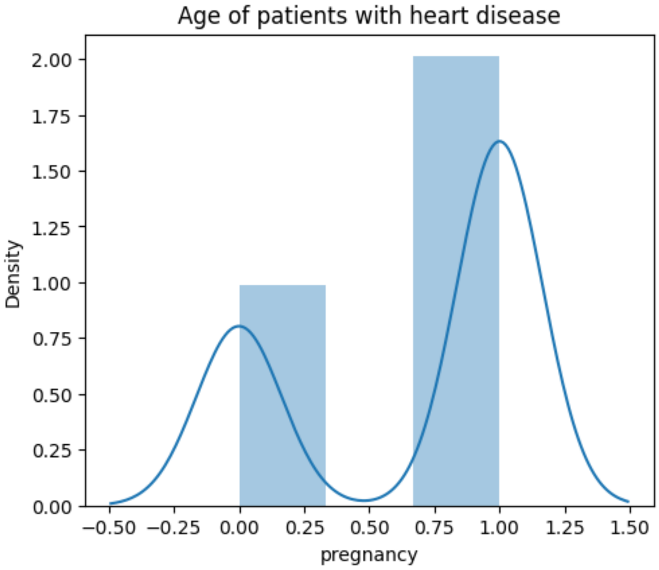

**Figure 12.** Pregnant Patients with Heart Disease

Figures 11 and 12 illustrate the distribution of the 'Pregnancy' variable for two distinct groups: patients without heart disease (Figure 4.10) and patients with heart disease (Figure 4.11). By plotting these distributions separately, it becomes possible to visually compare the two groups and identify any notable differences or emerging patterns. This comparison may reveal potential associations between pregnancy histories and the presence or absence of heart disease within the dataset (Khadidos et al., 2024).

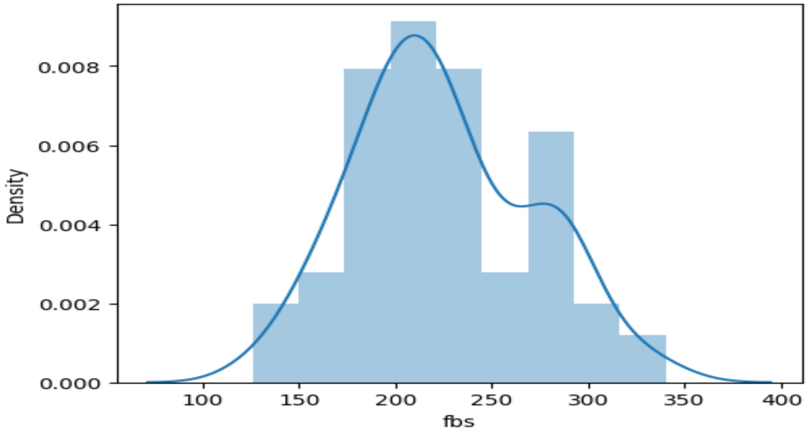

**Figure 13.** Fbs Patience without Heart Disease

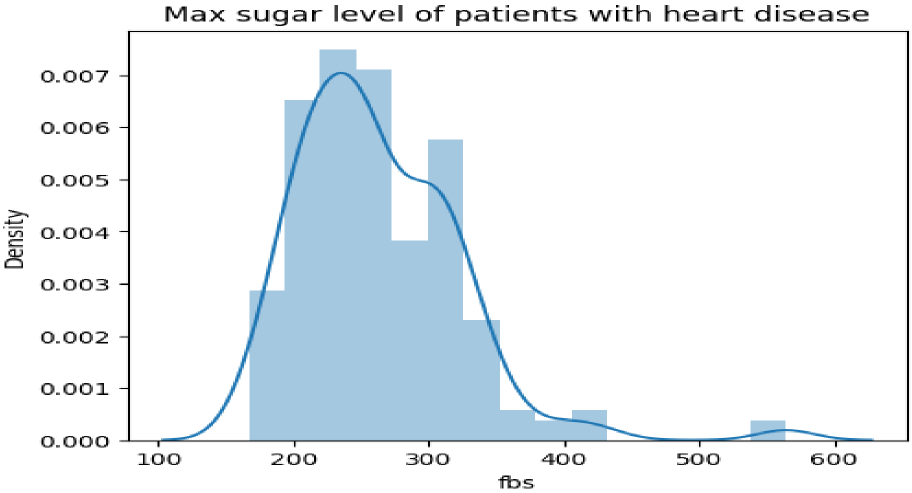

**Figure 14.** Fasting Blood Sugar in Patients with Heart Disease

Figures 13 and 14 present the distributions of the 'fbs' (fasting blood sugar) variable for patients without heart disease (Figure 13) and those with heart disease (Figure 14). By comparing these subplots, one can visually assess differences or patterns in fasting blood sugar levels between the two groups. This comparison may help identify potential associations between elevated fasting blood sugar and the presence of heart disease.

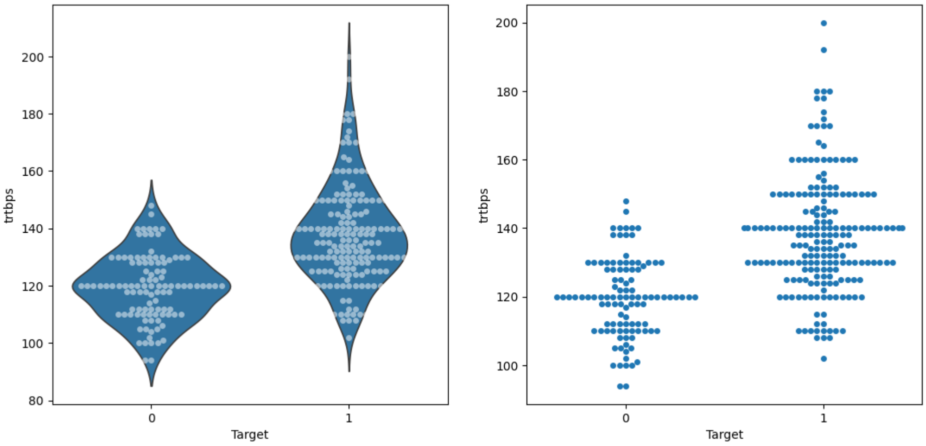

**Figure 15.** Relationship between Resting Blood Pressure ('trtbps') and Heart Disease

**Table 1.** Accuracy Achieved from the Evaluation Experiment

| Classificati | Purpose | Navie | Support | K-Nearet | Decision | Logistic |
|---|---|---|---|---|---|---|
| Accuracy | 86.89% | 78.69% | 77.05% | 83.61% | 75.41% | 85.25% |

Table 2 presents the accuracy results obtained from the experimental evaluation of different algorithms. It compares the performance of three classification algorithms used for predicting cardiovascular disease based on five selected attributes, alongside the proposed model. The proposed model achieved the highest accuracy at 86.89%, outperforming the other techniques. Among the standard algorithms, Logistic Regression demonstrated the closest performance to the proposed model. These findings suggest that further improvements in data preprocessing may enhance model tuning and overall performance (Huma et al., 2024).

## 5. Comparison between algorithms

A thorough evaluation of different classification algorithms was carried out to determine the most efficient method for predicting cardiovascular disease. The aim was to identify the algorithm offering the highest predictive accuracy for heart disease by analyzing their performance metrics.

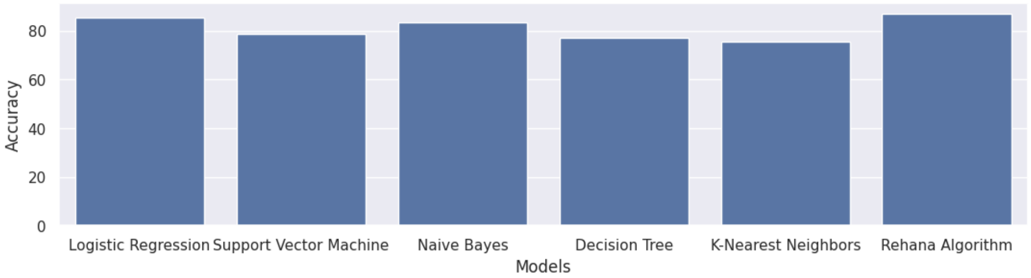

**Figure 16.** Accuracy Comparison of Algorithms

Figure 16 presents a bar chart illustrating the accuracy scores of various algorithms. The x-axis represents the names or labels of the algorithms, while the y-axis indicates their corresponding accuracy scores. The height of each bar reflects the performance of each algorithm. The chart highlights noticeable differences in accuracy among the evaluated models, with the proposed algorithm achieving the highest accuracy score.

## 6. Conclusion

This study explored the application of various classification algorithms—Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM)—for heart disease prediction. Using the Cleveland dataset from the UCI Machine Learning Repository, the data was split into 80% for testing and 20% for training to evaluate model performance on unseen data. After preprocessing to handle inconsistencies and missing values, SVM achieved the highest accuracy among the baseline models at 64.4%, suggesting it was most effective at capturing patterns within the data. Despite this, a key limitation was the inability to accurately identify early risk factors for heart disease, which is essential for early diagnosis and intervention. This shortcoming could be due to constraints within the dataset or the limited scope of input features. Identifying such risk factors is vital for preventative healthcare and improved patient outcomes. To address the limitations of traditional algorithms and improve performance, a proposed model combining the Random Forest (RF) algorithm and Random Search for Hyperparameters tuning was implemented. This model demonstrated superior performance, achieving an accuracy of 86.89%, significantly outperforming earlier models. The ensemble nature of Random Forest helps mitigate overfitting and enhances generalizability by aggregating results from multiple decision trees. The use of Random Search further optimized the model by efficiently identifying optimal Hyperparameters.

**References**

1. Hu N, Zhang Y, Wei Z, Yu R.” Associate between Renal Dysfunction Subtypes and Vertebral Fracture in Patients with type 2 Diabetes: A longitudinal study”. Diabetes, Metabolic Syndrome and Obesity 18:1915-1924. 11 June 2025.
2. Newson RS, Artime E, Bower J, Spaepen E, “Understanding Suboptimal Insulin Use in Type 1 and Type 2 Diabetes: A Cross- Sectional Survey of People with Diabetes”. Patient Preference and Adherence 2025. 19:1625-1638. 31 May 2025.
3. Chen R, Zhang K, Liu H, Liu L. “ Lipoprotein(a) as a Risk Factor for recurrent Ischemic Stroke in Type 2 Diabetes”. Diabetes, Metabolic Sndrome and Obesity 2025, 18:1631-1641. 17 May 2025.
4. Wu y, Li T, Chen b, Sun Y, Song L.” Tryptophan Indole Derivatives: key Players in type 2 Diabetes Mellitus “Diabetes, Metabolic Sndrome and Obesity 2025. 18:1551-1562. 10 May 2025.
5. Ute Schaefer Graf, Angela Napoli, "Diabetes in Pregnancy: a new decade of challenges ahead”. National Center for biotechnology Information.DOI:10.1007/s00125-018-4545-y. 22 Jan 2018.
6. Abbas Alam Choudhury, V. Devi Rajeswari.” Gestational Diabetes mellitus- A metabolic and reproductive disorder” Biomedicine & Pharmacotherap. Volume 143, November 2021. https://doi.org/10.1016/j.biopha.2021.112183.
7. Brandy Wicklow, Ravi Retnakaran. “ Gestational Diabetes Mellitus and its Implications across the life span”. Diabetes & Metabolic. 2023; 47(3):333-344. DOI: https://doi.org/10.4093/dmj.2022.0348.
8. Karolina Karcz, Barbara Krolak- olejnik. “ Impact of Gestational Diabetes mellitus on Fetal Growth and Nutritional Status in NewBorns”.Nutrients 2024, 16(23), 4093. 27 Nov 2024. https://doi.org/10.3390/nu16234093.
9. Keerthana G & Mohanasundaram R.” Early Prediction of diabetics using three layer classifier model and improves dimensionality reduction” Multi Media tools and application. 4 march 2025.
10. Souad Larabi marie- Sainte, Linah Abdurahmah. “ Curret Technique for Diabetes Predicition: Review and Case Study” MDPI. 29, Oct 2019.