

Heart Failure Prediction

Abstract

Accurate prediction of heart failure can help prevent life-threatening situations. Several factors contribute to the risk of heart failure, including underlying heart diseases such as coronary artery disease or heart attack, diabetes, hypertension, obesity, certain medications, and lifestyle habits such as smoking and excessive alcohol intake. Machine learning approaches to predict and detect heart disease hold significant potential for clinical utility but face several challenges in their development and implementation. This research proposes a machine learning metamodel for predicting a patient's heart failure based on clinical test data. CAD (coronary artery disease) is one of the most common types of heart disease. In the medical field predicting the heart disease has become a very complicated and challenging task, requires patient previous health records and in some cases, they even need Genetic information as well. So, in this contemporary lifestyle there is an urgent need of a system which will predict accurately the possibility getting heart disease. Predicting a heart disease in early stage will save many people's Life. There were many heart disease prediction systems available at present, the Authors have been researched well and proposed different Classification and prediction algorithms but each one has its own limitations. The main objective of this paper is to overcome the limitations and to design a robust system which works efficiently and will be able to predict the possibility of heart failure accurately. This paper uses the data set from the Kaggle repository. This work is implemented using many algorithms such as Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, Gaussian Naïve Bayes. The metamodel is trained and tested utilizing a combined dataset comprising five well-known heart datasets (Statlog Heart, Cleveland, Hungarian, Switzerland, and Long Beach), all sharing 11 standard features.

1. Introduction

Heart failure is a complex and potentially life-threatening condition that significantly burdens healthcare systems worldwide. It is a pathophysiologic condition in which the heart's inability to pump blood at a rate sufficient to meet the needs of the body's metabolizing tissues results from faulty cardiac function [1,2]. It includes a few heart-related illnesses, such as coronary artery disease, heart attacks, heart failure, arrhythmias, and several other cardiovascular ailments. Heart disease is a leading cause of death globally [3], accounting for many premature deaths and posing a significant burden on healthcare systems. Heart disease is a common and significant health issue in many parts of the world [4]. The American Heart Association says that heart failure is projected to increase dramatically [5]. Accurate prediction of heart failure can play a vital role in early detection and prevention of adverse outcomes, ultimately leading to improved patient outcomes and reduced healthcare costs.

Timely and accurate detection of heart failure is crucial for effective management and treatment [6]. Detecting heart failure early allows for prompt intervention and the implementation of appropriate medical strategies, which can help slow the progression of the disease, alleviate symptoms, and improve the patient's quality of life. Early detection can also reduce the risk of complications and hospitalizations associated with advanced stages of heart failure. From 1989 [7] to now, there have been many approaches to finding the best methods for cardiac failure prediction. In 2017, Simge et al. [8] used Matlab and WEKA to find the best way to detect heart failure disease and obtained a good accuracy of 67.7% for the ensemble subspace discriminant algorithm and the decision tree algorithm. Then, in 2018, Ali et al. [9] utilized the Cleveland dataset [10] for their studies and obtained 84% accuracy for the Naive Bayes algorithm. Further, in 2019, Saba et al. [11] performed prediction for heart diseases and obtained 84.85% accuracy for the logistic regression (SVM) technique. However, most of them used the same dataset from the UCI repository [12] which contains 300 records. This is a rather limited amount of data for machine learning training.

Machine learning techniques have drawn a lot of attention in the medical field lately because of their potential to help with the detection and prediction of cardiac disease [13]. Large volumes of clinical data may be analyzed by machine learning algorithms to find links and patterns that are not immediately obvious to human practitioners [14]. These algorithms can harness the power of computer models to make accurate predictions and provide valuable insights into disease risk assessment. However, the development and implementation of machine learning models for heart failure prediction face several challenges [15]. The complexity of the cardiovascular system and the multifactorial nature of heart failure necessitate integrating diverse data sources, including clinical test data, medical imaging, and patient demographics. Data quality, feature selection, and model performance issues must be addressed to ensure reliable and clinically relevant predictions.

This research addresses these challenges by proposing a machine learning metamodel for predicting heart failure based on clinical test data. The metamodel incorporates several established machine learning algorithms, namely the Logistic Regression, SVM, Gaussian Naive Bayes (GNB), Random Forest (RF), Decision Tree (DT) to leverage their individual strengths in classification tasks. Combining these models into a metamodel aims to enhance predictive accuracy and robustness while reducing potential biases associated with individual algorithms. To evaluate the performance of the proposed metamodel, a combined dataset comprising five well-known heart datasets, including the Statlog Heart, Cleveland, Hungarian, Switzerland, and Long Beach datasets, is utilized. These datasets share 11 standard features such as age, chest pain type, sex, resting BP, fasting BS, cholesterol, resting ECG, exercise angina, maxHR, oldpeak, and ST-slope, widely used in previous heart disease prediction studies. By leveraging a diverse set of data sources, the metamodel aims to represent the underlying factors contributing to heart failure comprehensively.

The overall contribution of this paper includes the following:

- Integrate common machine learning algorithms such as Logistic Regression, SVM, Random Forest, Gaussian Naive Bayes, Decision Tree, and k-Nearest Neighbor into a metamodel framework, leveraging their strengths to improve predictive accuracy and model robustness.
- We have used a combined dataset from five different and well-known cardiac datasets, including Statlog Heart, Cleveland, Hungary, Switzerland, and Long Beach, to ensure a comprehensive representation of patient characteristics, clinical features, and risk factors and improve the metamodel's generalizability and applicability.

2. Research Design

2.1 Dataset Description

This dataset was created in September 2021 by FEDESORIANO, by combining five independent heart disease datasets with over 11 common features. The five datasets that have been used as per below:

- Cleveland [Year 1990]: 303 observations
- Hungarian [Year 1990]: 294 observations
- Switzerland [Year 1989]: 123 observations
- Long Beach VA [Year 1989]: 200 observations
- Stalog (Heart) Data Set [Year 1990]: 270 observations

Total: 1190 observations

Duplicated: 272 observations

Final dataset: 918 observations

This data set consists of 918 patient records and 11 significant attributes which plays key role in identifying the presence or absence of the heart disease. These 11 clinical attributes are preprocessed and trained to predict that there is the presence or absence of the heart disease.

Attribute	Description
Age	Age of the patient in years
Sex	Sex of the patient <ul style="list-style-type: none">• Male (M)• Female (F)
ChestPainType	Chest pain type <ul style="list-style-type: none">• TA: Typical Angina• ATA: Atypical Angina• NAP: Non-Anginal Pain• ASY: Asymptomatic
RestingBP	Resting blood pressure in mm Hg
Cholesterol	Serum cholesterol in mm/dl
FastingBS	Fasting blood sugar <ul style="list-style-type: none">• 1: If FastingBS > 120 mg/dl• 0: Otherwise
RestingECG:	Resting electrocardiogram results <ul style="list-style-type: none">• Normal: Normal• ST: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)• LVH: Showing probable or definite left ventricular hypertrophy by• Estes' criteria]
MaxHR	Maximum heart rate achieved (Numeric value between 60 and 202.
ExerciseAngina	Exercise-induced angina Y: Yes N: No
Oldpeak	Oldpeak = ST (Numeric value measured in depression)
ST_Slope	The slope of the peak exercise ST segment <ul style="list-style-type: none">• Up: Upsloping• Flat: Flat• Down: Downsloping

HeartDisease	Output/target class
	<ul style="list-style-type: none"> • 1: Heart disease • 0: Normal

Table 1. Dataset Attributes

2.2 Dataset Preprocessing

Data Pre-processing is one of the most important data mining tasks which includes preparation and transformation of data into suitable form for mining procedure. Data pre-processing reduces the size of the data and includes all techniques like data cleaning, integration relation between data and normalization of data. It also removes outliers and extracts features from data. It also involves removing feature noise from the data and suitable filling of missing features.

2.2.1 Data Cleaning

In data cleaning, missing (null) values, duplication of observations and outliers will be investigated.

- There are no null values in dataset.
- There is no duplicate data in the dataset.
- The categorical variables Sex, ChestPainType, FastingBS, RestingECG, ExerciseAngina, ST_Slope are transformed into factors.
- Handling of outliers will only be performed continuous data only, i.e. Age, RestingBP, Cholesterol, MaxHR and Oldpeak. Box plot is employed to identify the potential outliers visually.

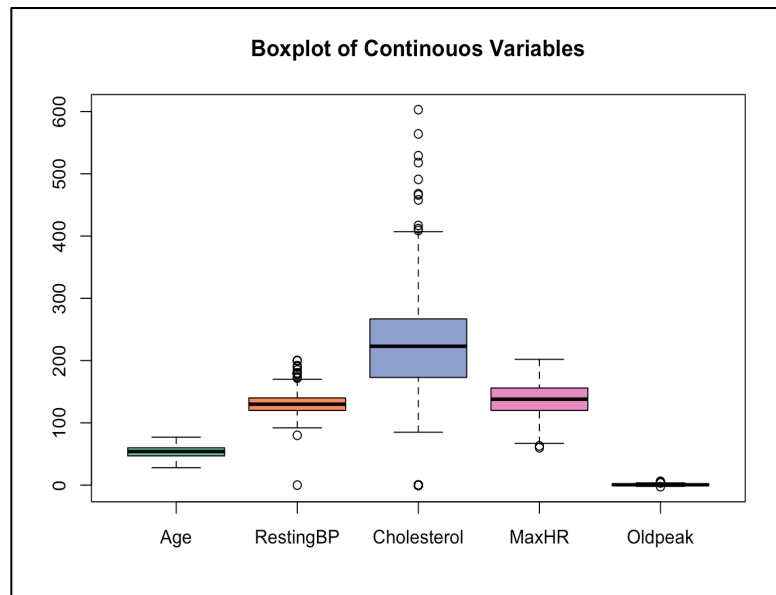


Fig 1. Boxplot to handle outliers.

From boxplots, the RestingBP and Cholesterol are having a few observations that are exceed that maximum range, while for RestingBP, Cholesterol and MaxHR are observed to have outliers that stay at below minimum range.

After removing the outliers for all continuous data columns, the dimension of dataset is updated from 918 rows to 702 rows with 12 variables. 216 rows of observations have been recognized as outliers and removed.

2.3 Feature Selection

Feature selection is a technique to select the best set of features (subset of relevant features) for use in construction of optimized models.

```

> feature_select$finalDecision
  Age      Sex ChestPainType RestingBP Cholesterol FastingBS RestingECG MaxHR ExerciseAngina Oldpeak
Confirmed Confirmed Confirmed Confirmed Rejected Confirmed Confirmed Confirmed Confirmed Confirmed
ST_Slope
Confirmed
Levels: Tentative Confirmed Rejected
>

```

Fig 2. Feature Selection.

According to the Boruta algorithm's output, the Cholesterol has been rejected. However, it is worth noting that high cholesterol levels have been scientifically linked to the development of fatty deposits in blood vessels, leading to reduced blood flow and an increased risk of heart disease. Considering this significant association, it is believed that Cholesterol may still hold valuable insights. Therefore, this variable will be retained for further analysis, i.e. to investigate possibilities of transforming the variable or combining it with other variables to create more informative features.

For all other variables in the dataset, the result shows that they are in *Confirmed* status, indicating that the features have been confirmed important by the Boruta algorithm and significantly contributes to predict the response variable (HeartDisease).

2.4 Data Exploration

Data exploration is essential to gain a better understanding of the dataset. It involves examining and visualizing the data to identify patterns, trends, relationships, and anomalies with the aim of extracting useful insights and inform further analysis or decision-making.

2.4.1 Univariate Analysis

Analysis carried out by considering one variable at a time.

2.4.1.1 Histogram

All continuous data is analyzed with histogram to observe the overall distribution.

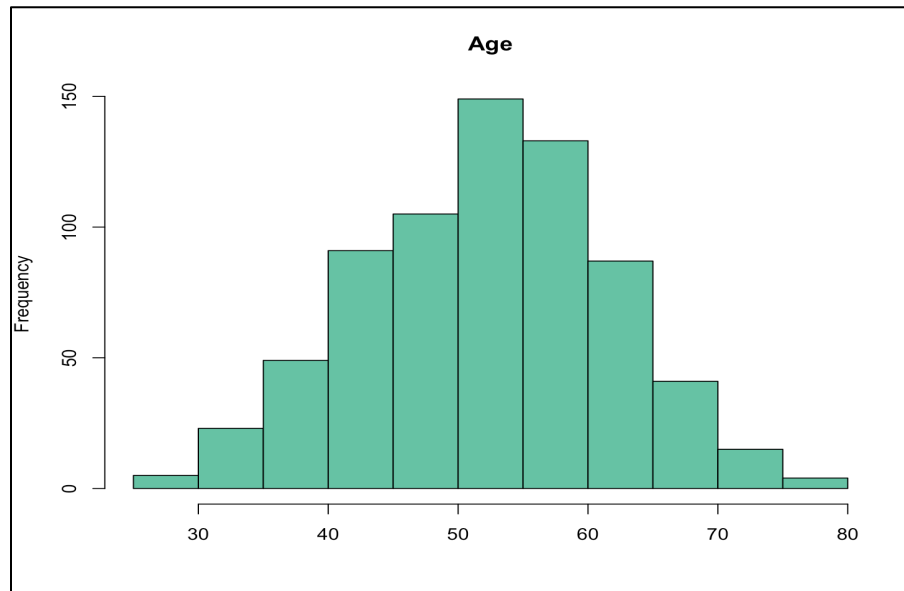


Fig 3. Histogram of Age

From above histogram of Age, it is observed to be normally distributed, no outlier, skewed data, or any abnormality.

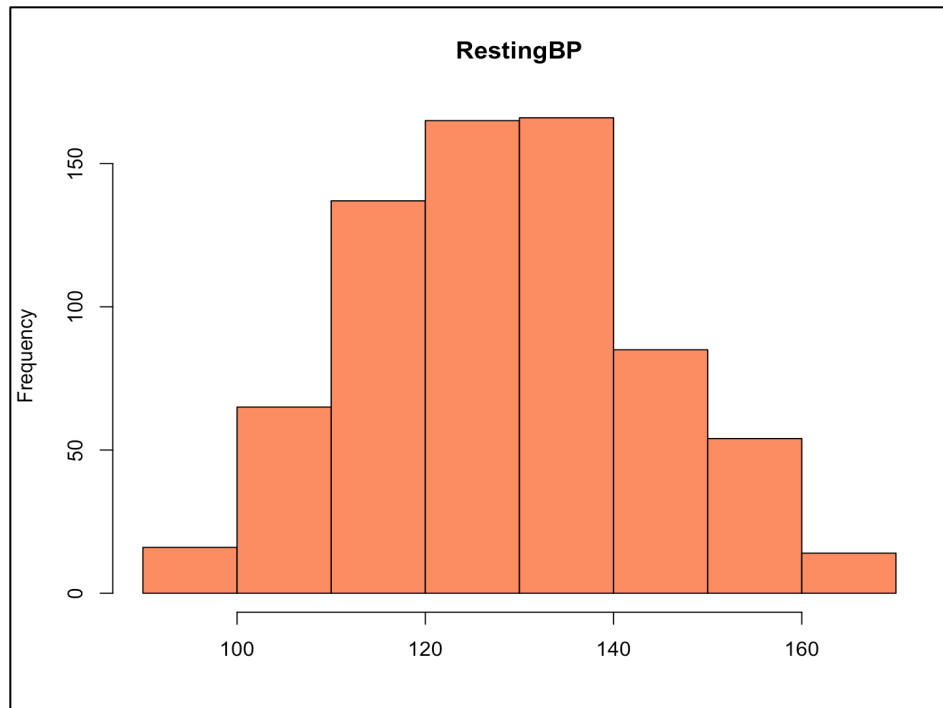


Fig 4. Histogram of RestingBP

From above histogram of RestingBP, it is observed to be normally distributed, no outlier, skewed data, or any abnormality.

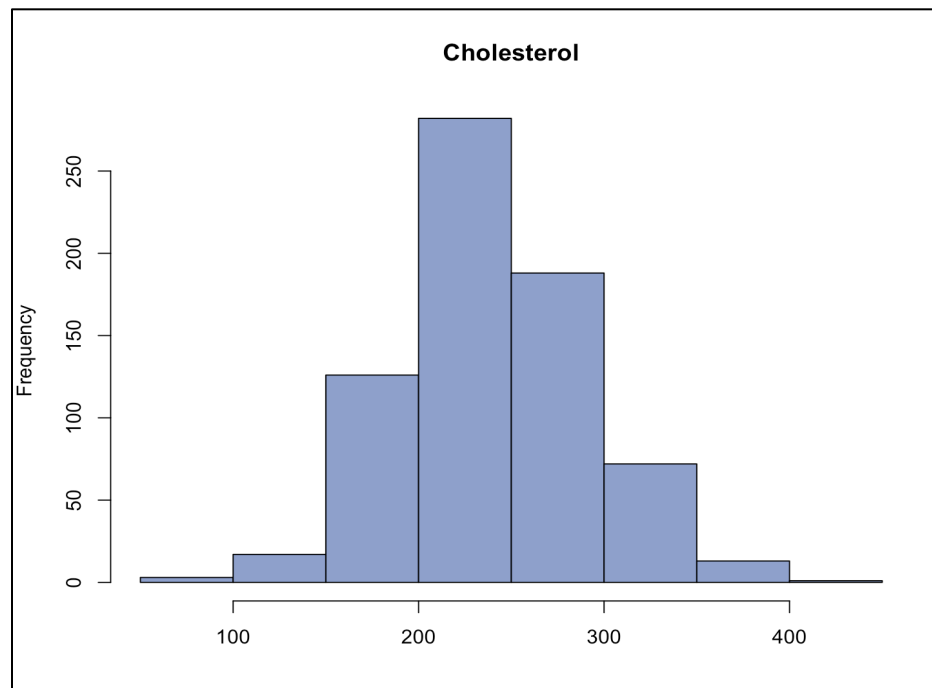


Fig 5. Histogram of Cholesterol

From above histogram of Cholesterol, it is observed to be normally distributed, no outlier, skewed data, or any abnormality.

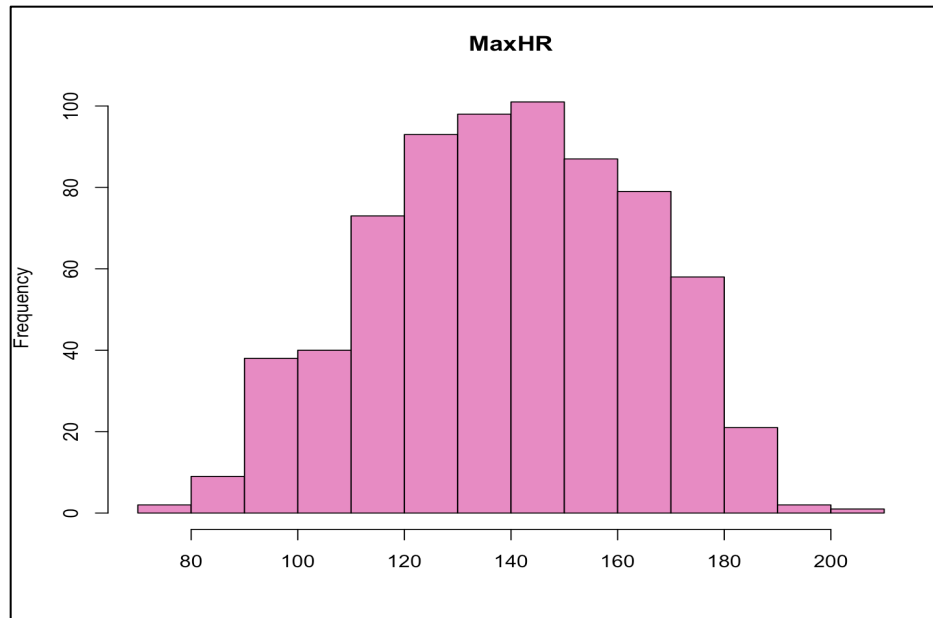


Fig 6. Histogram of MaxHR

From above histogram of MaxHR, it is observed to be normally distributed, no outlier, skewed data, or any abnormality.

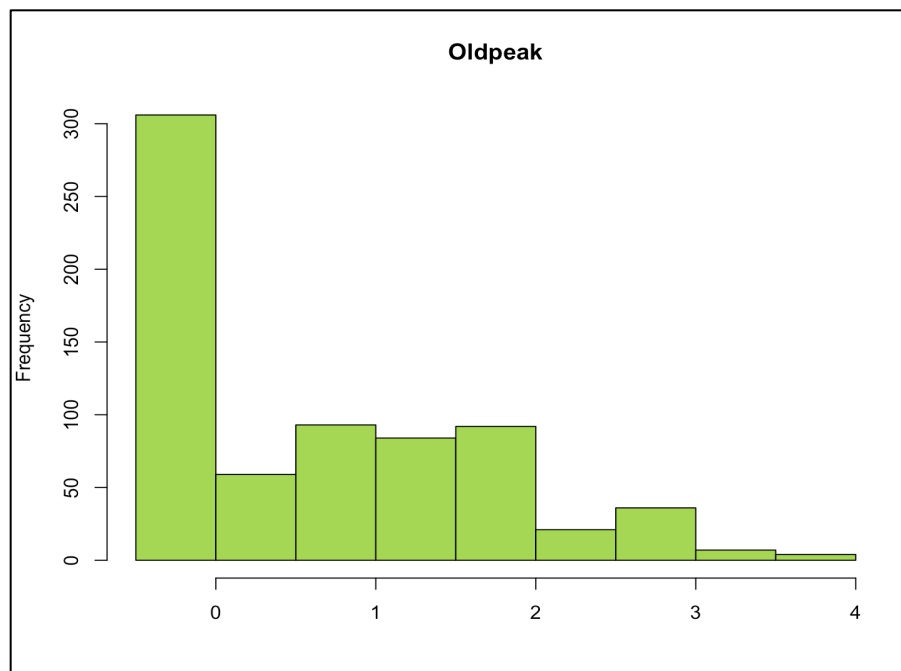


Fig 7. Histogram of Oldpeak

For Oldpeak, data seems skewed to the left. Further inspect the raw (before transformed) data, there are around 43% of observation has 0 as the Oldpeak value, suggests that there is no ST segment depression observed in the electrocardiogram (ECG) reading (Oldpeak = ST (Numeric value measured in depression)).

2.4.2 Bivariate Analysis

The relationships between pairs of variables are explored.

2.4.2.1 Correlation and Heatmap

The relationship between 2 continuous variables (normally distributed) are investigated.

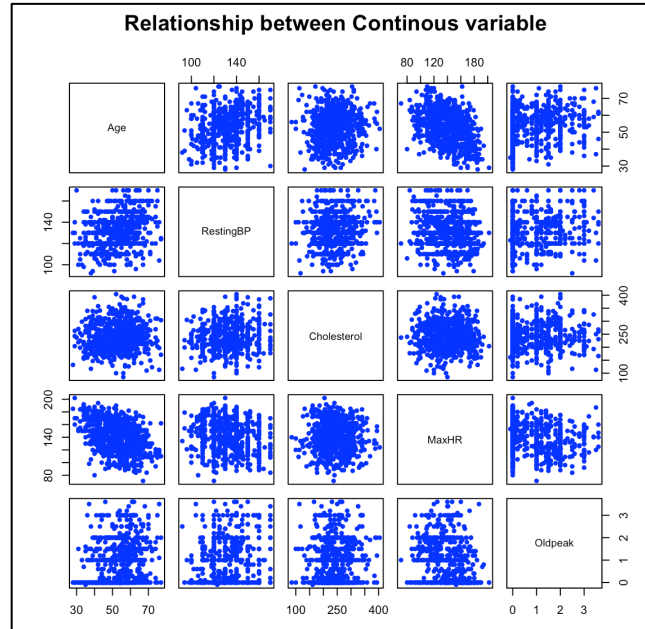


Fig 8. Correlation matrix

The Correlation Matrix plots indicate weak relationships between the variables. The data points are widely scattered, showing no clear trend or correlation between pairs such as Age and MaxHR, or Cholesterol and RestingBP. This suggests that if there are relationships, they may be non-linear or influenced by other factors not included in these variables.

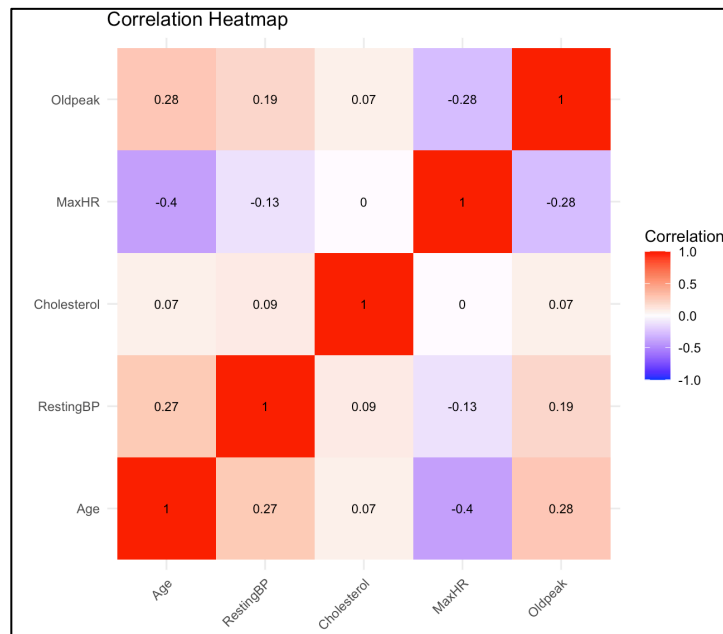


Fig 9. Correlation Heatmap

The heatmap suggests that Age has a moderate positive correlation with RestingBP and Oldpeak, and a moderate negative correlation with MaxHR. RestingBP and Oldpeak also show a slight positive correlation. There are no strong correlations between Cholesterol and any of the other variables. Overall, the heatmap indicates some expected relationships between age, heart rate, and blood pressure, but cholesterol levels appear to be largely independent of the other variables in this dataset.

2.4.2.2 Barchart

Bar charts for categorical variable (Sex, ChestPainType, ExerciseAngina, RestingECG, ST_Slope, FastingBS) are generated to see the distribution of target variable, *heart disease*, across each category within the independent variable.

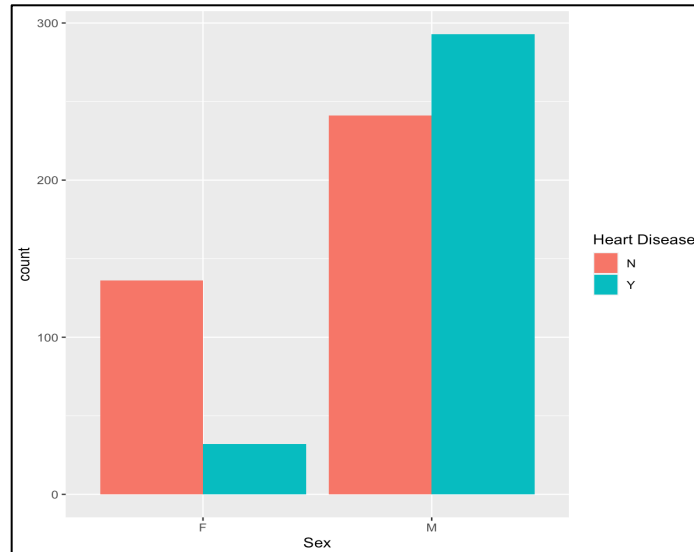


Fig 10. Bar chart for Sex with HeartDisease

By above bar chart it is observed, most of the heart disease individuals is male.

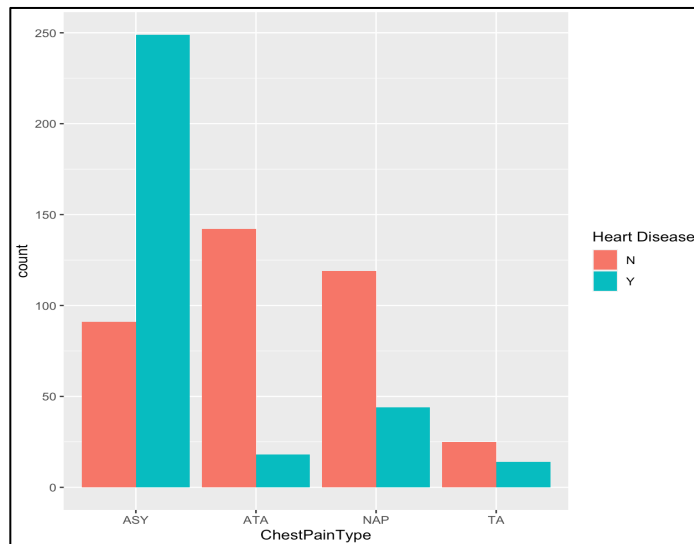


Fig 11. Bar chart for ChestPainType with HeartDisease

By above bar chart it is observed, most of the heart disease individuals are with Asymptomatic Chest Pain Type.

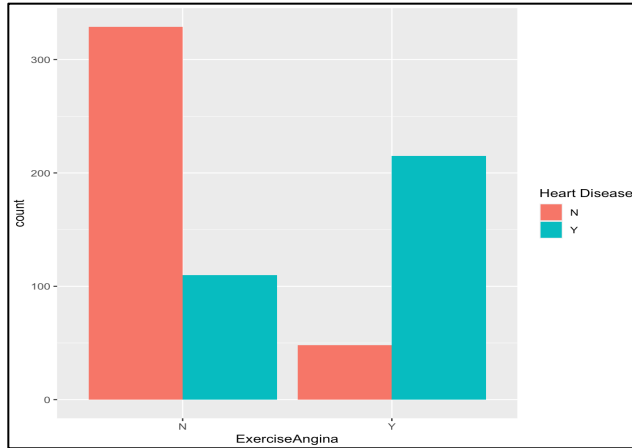


Fig 12. Bar chart for ExerciseAngina with HeartDisease

By above bar chart it is observed, most of the non-heart disease individuals without exercise-induced angina.

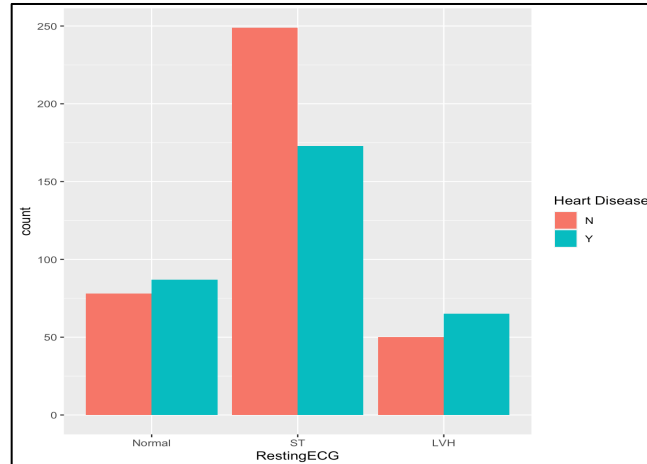


Fig 13. Bar chart for RestingECG with HeartDisease

Individuals who have ST resting electrocardiogram have the highest percentage of non-heart disease.

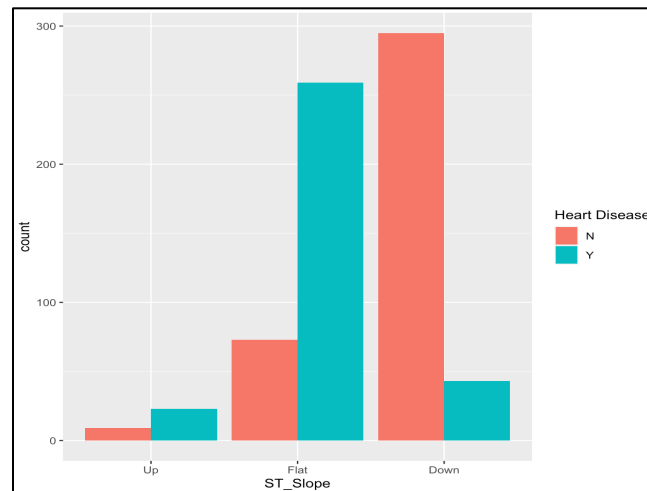


Fig 14. Barchart for ST_Slope with HeartDisease

Most of the heart disease individuals with Flat ST slope while most of the non-heart disease individuals with Down ST slope.

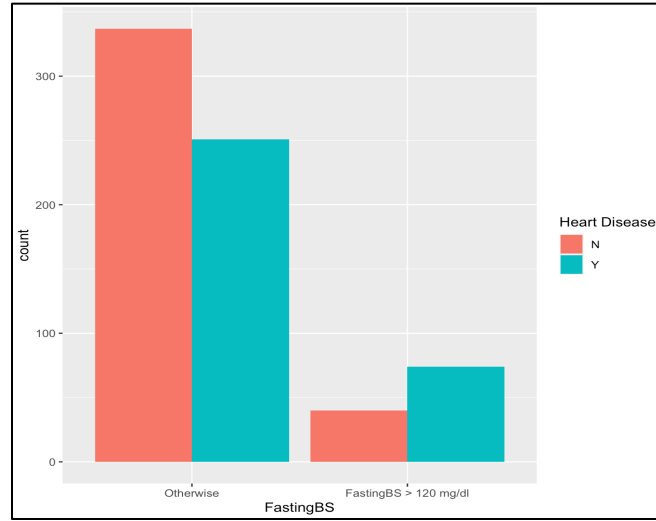


Fig 15. Bar chart for FastingBS with HeartDisease

For individuals with fasting Otherwise has the highest percentage of non-heart disease.

2.4.3 Multivariate Analysis

The relationships between multiple variables are studied simultaneously.

2.4.3.1 Scatterplots

Scatterplots for continuous variable (Age, RestingBP, Cholesterol, MaxHR) are generated to see the relationships between them with the consideration of heart disease, the target variable as factor.

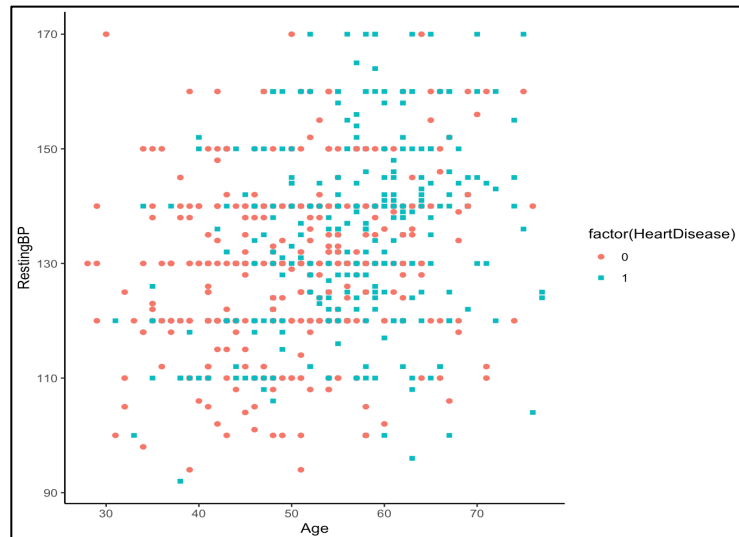


Fig 16. Scatterplot between Age and RestingBP

The scatterplot shows that as age and resting blood pressure increases, there is an increase in the proportion of individuals with heart disease compared to those without heart disease. This suggests a positive correlation between resting blood pressure and heart disease.

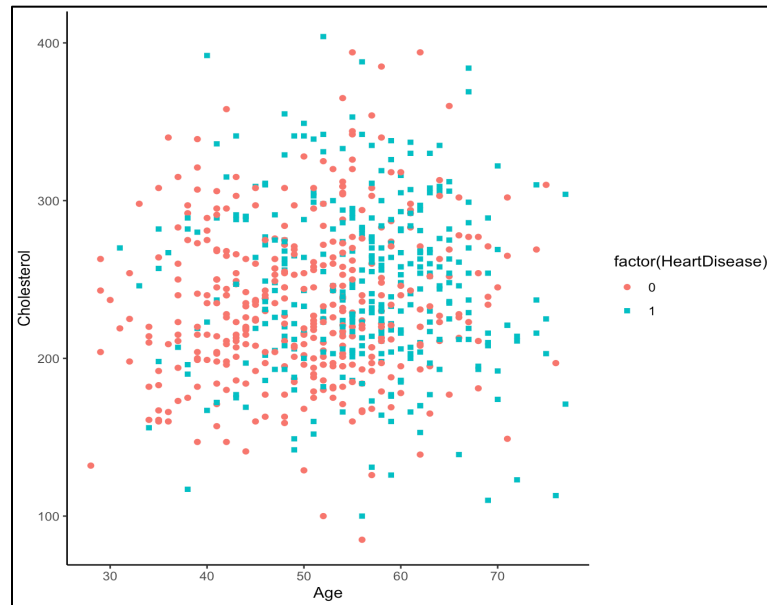


Fig 17. Scatterplot between Age and Cholesterol

The scatterplot shows that as age and cholesterol level increases, there is a slight increase in the proportion of individuals with heart disease compared to those without heart disease. This suggests a positive correlation between an increase in age, cholesterol with the presence of heart disease.

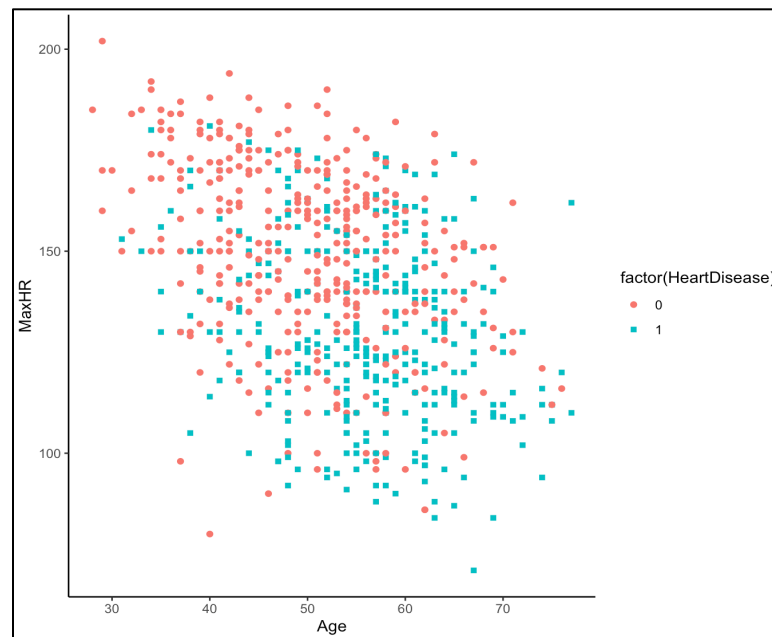


Fig 18. Scatterplot between Age and MaxHR

The scatterplot shows that when the maximum heart rate decreases with an increase in age, there is an increase in the proportion of individuals with heart disease. This suggests a negative correlation between maximum heart rate and heart disease.

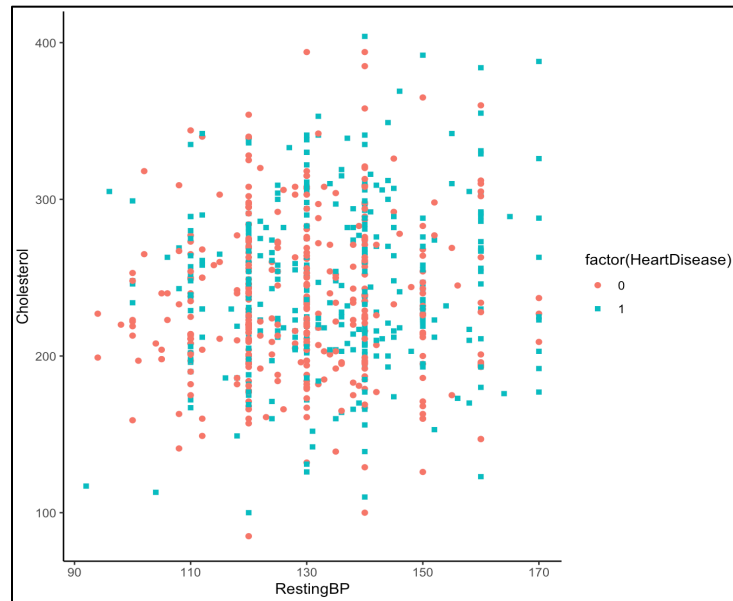


Fig 19. Scatterplot between RestingBP and Cholesterol

The scatterplot shows that as cholesterol levels and resting blood pressure increases, there is an increase in the proportion of individuals with heart disease. This suggests a positive correlation between cholesterol levels, resting blood pressure and heart disease.

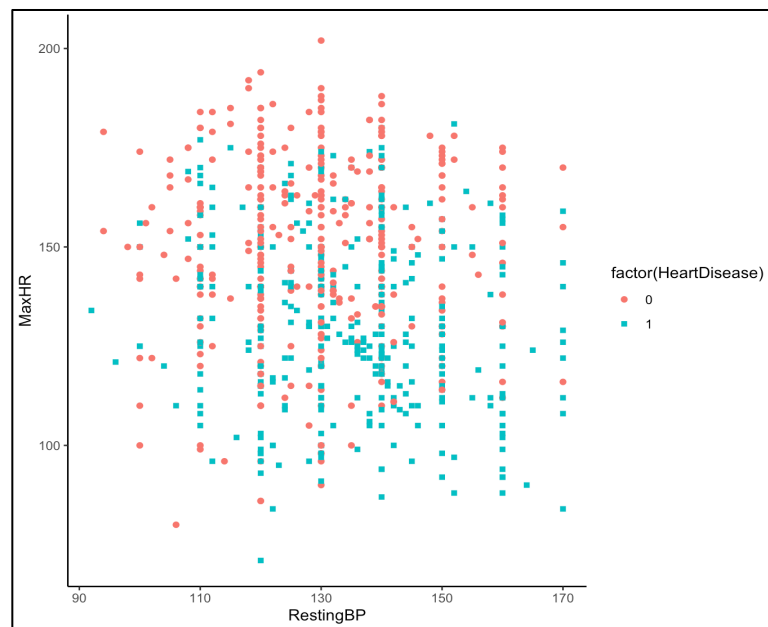


Fig 20. Scatterplot between RestingBP and MaxHR

The scatterplot shows the relationship between resting blood pressure and maximum heart rate, with individuals with no heart disease (red dots) and those with heart disease (blue dots) represented separately. The scatterplot suggests that there is no clear pattern or correlation between resting blood pressure and maximum heart rate for both groups.

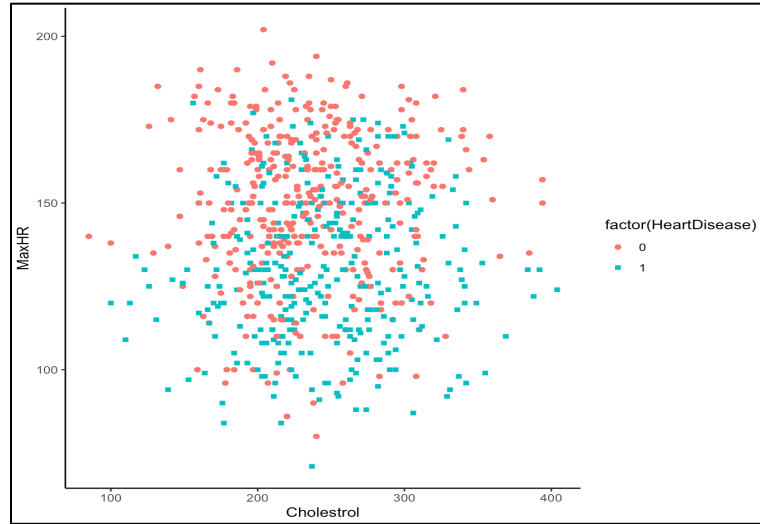


Fig 21. Scatterplot between Cholesterol and MaxHR

The scatterplot visually represents the correlation between maximum heart rate and cholesterol levels, with two distinct groups: individuals without heart disease (represented by red dots) and those with heart disease (represented by blue dots). The plot clearly demonstrates that as cholesterol levels increase, maximum heart rate decreases, indicating a heightened risk of heart disease.

2.5 Data Analysis and Model Development

Data modeling refers to the process of developing a predictive model based on available data and problem requirements. Appropriate modeling algorithms are selected to train the model, with the aim of producing a reliable and accurate model that capable of generating valuable insights.

Modeling algorithms used in this project includes:

1. Logistic Regression
2. Support Vector Machine
3. Decision Tree
4. Random Forest
5. Gaussian Naïve Bayes

2.6 Model Evaluation

Model evaluation is the process of assessing the performance and effectiveness of a trained model using various evaluation metrics. It involves measuring how well the model generalizes to unseen data and how accurately it predicts or describes the target variable. In this case, confusion matrix, as generated in the model development phase, is extremely useful for the evaluation metrics' calculation.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig 22. Confusion Matrix Indicator

In the process of deciding the best model, various evaluation metrics specifically for classification problem are considered to assess the performance of the models, such as:

1. **Accuracy**: This measures the overall correctness of the model and is calculated as the number of correct predictions divided by the total number of predictions.

$$\text{Formula: Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

2. **Balanced Accuracy**: This is used to handle imbalanced datasets and is the average of recall obtained on each class.

$$\text{Formula: Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

3. **Sensitivity (also known as Recall or True Positive Rate)**: To measures the model's ability to correctly identify positive instances, or in other word the ability to avoid false negative errors.

$$\text{Formula: Sensitivity} = \frac{TP}{TP+FN}$$

4. **Specificity (also known as True Negative Rate)**: This measures the proportion of actual negatives that are correctly identified.

$$\text{Formula: Specificity} = \frac{TN}{TN+FP}$$

5. **Precision (also known as Positive Predictive Value)**: This measures the proportion of positive identifications that were correct.

$$\text{Formula: Precision} = \frac{TP}{TP+FP}$$

3. Model Development

3.1 Logistic Regression

	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Original	90.07	90.12	89.19	91.04	91.76
Over-sampled	90.78	91	86.49	95.52	95.52
Under-sampled	88.65	88.77	86.49	91.04	91.43
Combination	89.36	89.37	89.19	89.55	90.41

Table 2. Evaluation Metrics of Logistic Regression

The 'Over-sampled' method yields the highest Accuracy (90.78%) and Balanced Accuracy (91%), indicating it is the most reliable for overall performance. It also has the highest Specificity and Precision, suggesting it is particularly strong at predicting the true negatives and the proportion of positive identifications that were correct.

However, its Sensitivity is lower compared to the 'Original' and 'Combination' methods, which means it is not as good at identifying true positives. The 'Under-sampled' method shows the lowest performance across all metrics, suggesting it may be the least effective at handling the dataset among the methods compared. Considering these factors, the 'Over-sampled' model appears to offer the best balance between all evaluated metrics, making it the preferred model.

3.2 Support Vector Machine (SVM)

	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Original	83.69	83.61	87.50	79.71	81.82
Over-sampled	82.27	82.97	90.62	75.32	75.32
Under-sampled	83.69	83.61	87.50	79.71	81.82
Combination	84.4	84.37	88.73	80	81.82

Table 3. Evaluation Metrics of SVM

The 'Combination' method exhibits the highest Accuracy (84.4%) and Balanced Accuracy (84.37%), suggesting it performs best in terms of overall correct predictions and balanced performance across classes. This method also has a good balance between Sensitivity and Specificity, indicating it is relatively effective at identifying true positives and true negatives.

The 'Over-sampled' method, while having the highest Sensitivity (90.62%), suffers in Specificity (75.32%) and Precision (75.32%), indicating it might be predicting more false positives. Both the 'Original' and 'Under-sampled' methods show identical metrics, with moderate performance across all categories. Therefore, the 'Combination' method would be the best choice among the evaluated models.

3.3 Decision Tree

	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Original	77.3	77.39	77.01	77.78	84.81
Over-sampled	77.3	77.75	75.86	79.63	85.71
Under-sampled	77.3	77.39	77.01	77.78	84.81
Combination	75.89	76.25	74.71	77.78	84.42

Table 4. Evaluation Metrics of Decision Tree

The performance of the Decision Tree model is relatively consistent across different sampling methods when looking at Accuracy, which remains the same for the original, over-sampled, and under-sampled methods at 77.3%. However, the 'Over-sampled' method shows a slight improvement in Balanced Accuracy (77.75%) and a notable increase in Specificity (79.63%) and Precision (85.71%) compared to the 'Original' and 'Under-sampled' methods, which have identical metrics. Sensitivity is slightly lower for the 'Over-sampled' method than the others, but the difference is not substantial. The 'Combination' method shows a decrease in all metrics compared to the other methods, suggesting that this method may not be as effective.

Given these observations, the 'Over-sampled' method appears to offer a marginal improvement in Precision and Specificity without a loss in Accuracy. This could suggest that the 'Over-sampled' model may be slightly better at correctly predicting the positive class and at producing a higher proportion of relevant results.

Original

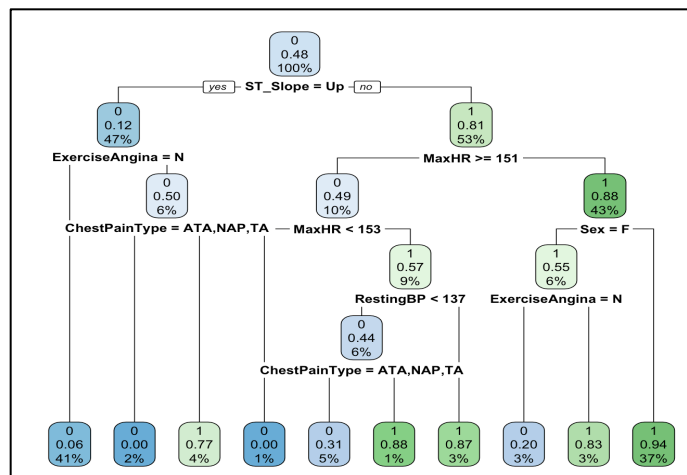


Fig 23. Decision Tree for Original dataset

Oversample

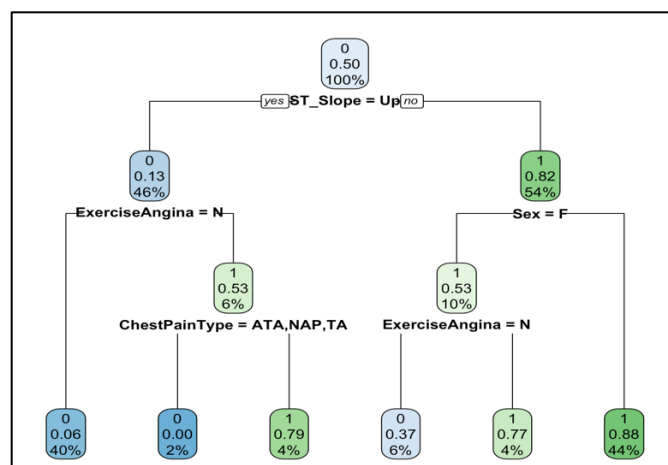


Fig 24. Decision Tree for Oversample dataset

Undersample

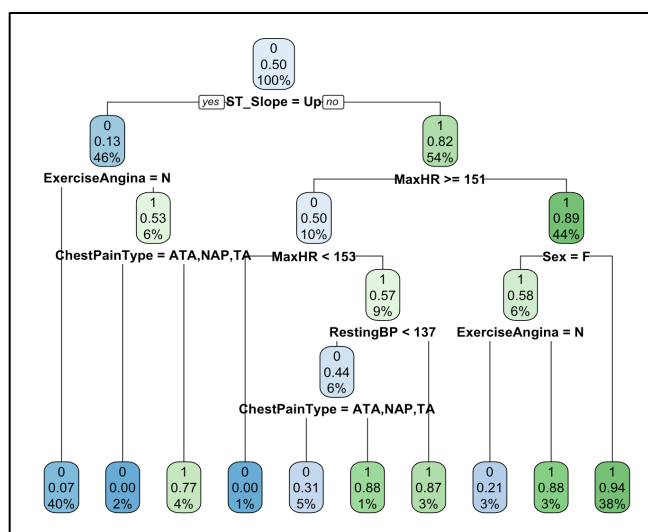


Fig 25. Decision Tree for Undersample dataset

Combination

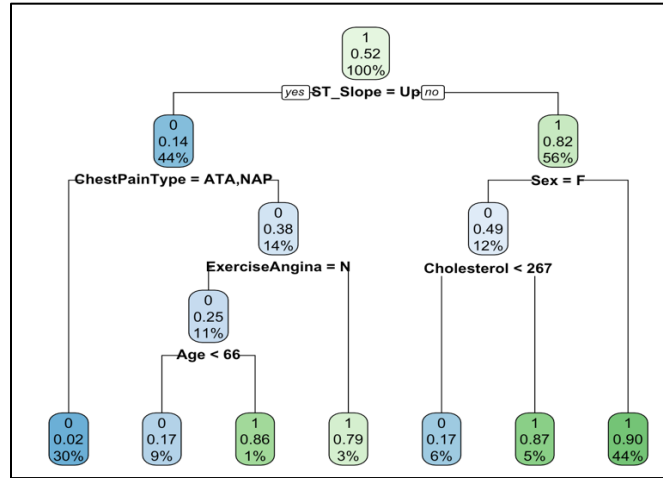


Fig 26. Decision Tree for Combination dataset

3.4 Random Forest

	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Original	86.52	86.24	90.67	81.82	85
Over-sampled	85.11	84.64	92	77.27	82.14
Under-sampled	89.36	89.36	89.33	89.39	90.54
Combination	85.11	84.91	88	81.82	84.62

Table 5. Evaluation Metrics of Random Forest

The 'Under-sampled' method shows the highest Accuracy (89.36%) and Balanced Accuracy (89.36%), indicating it is the best performing technique in terms of overall correct predictions balanced across classes for the Random Forest model. It also has the highest Specificity (89.39%) and Precision (90.54%), suggesting it is particularly effective at predicting true negatives and the proportion of positive identifications that are correct. Although the Sensitivity (89.33%) is slightly lower than the 'Over-sampled' method (92%), the 'Under-sampled' method still maintains a high rate of correctly identifying true positives. On the other hand, the 'Over-sampled' method, despite its highest Sensitivity, has the lowest scores in both Specificity and Precision, which might indicate a tendency to overpredict the positive class, leading to more false positives. The 'Original' and 'Combination' methods show moderate performance across all metrics.

Considering the significant improvement in Accuracy, Balanced Accuracy, Specificity, and Precision, the 'Under-sampled' stands out as the best performer among the different sampling techniques.

3.5 Gaussian Naïve Bayes

To apply the Gaussian Naïve Bayes model, it's necessary to address the distribution of the 'Oldpeak' variable, which is not normally distributed. To transform 'Oldpeak' into a normal distribution, logarithmic transformation is applied to its values. In cases where the log transformation results in negative infinity ("-Inf") values, these are substituted with 0. Below is the histogram illustrating the distribution of 'Oldpeak' after these adjustments have been made.

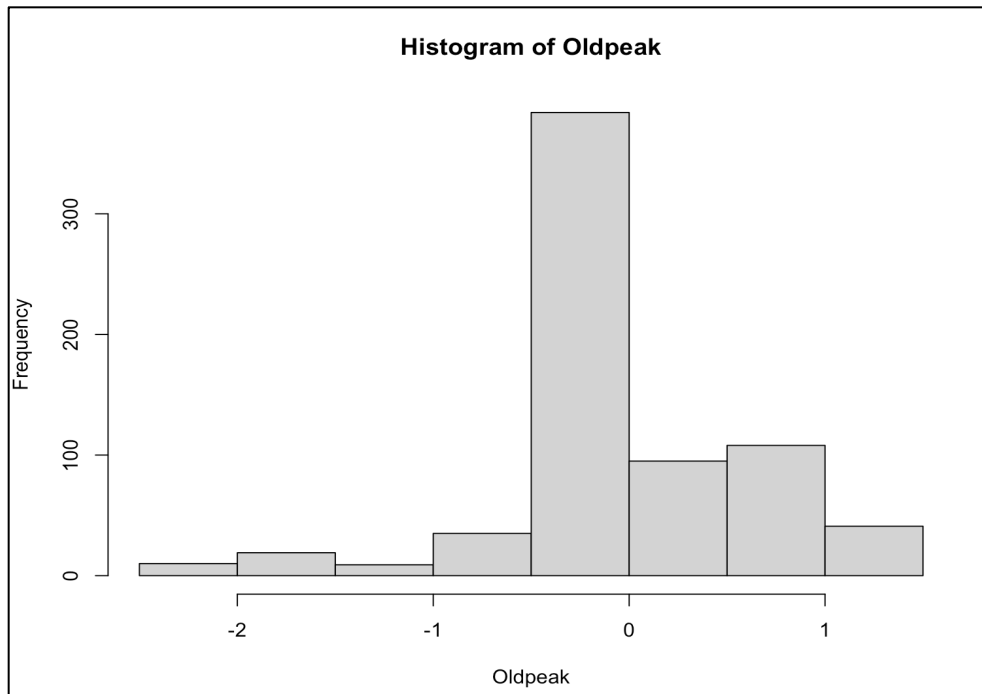


Fig 24. Histogram of Oldpeak after modifications

	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Original	85.11	85.03	86.49	83.58	85.33
Over-sampled	85.82	85.78	86.49	85.07	86.49
Under-sampled	85.82	85.78	86.49	85.07	86.49
Combination	84.4	84.43	83.78	85.07	86.11

Table 6. Evaluation Metrics of Gaussian Naïve Bayes

The 'Over-sampled' and 'Under-sampled' methods yield identical improvements in Accuracy and Balanced Accuracy over the 'Original' method, both scoring 85.82% and 85.78% respectively. This indicates that both methods are equally effective in overall correct predictions and in achieving a balance across classes for the Gaussian Naïve Bayes model. They also maintain the same Sensitivity (86.49%) and Specificity (85.07%), with an increase in Precision to 86.49% compared to the 'Original' method. This suggests that when the data is either over-sampled or under-sampled, the model is slightly more precise in its positive predictions.

The 'Combination' method shows a decrease in most metrics compared to the 'Original', 'Over-sampled', and 'Under-sampled' methods, suggesting it may not be as effective. Considering the improvement in Precision without a loss in other metrics, the 'Over-sampled' and 'Under-sampled' models appear to be the best performing among the four approaches.

3.6 Models Evaluation

Model	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision
Logistic Regression	90.07	91	86.49	95.52	95.52
Support Vector Machine	84.4	84.37	88.73	80	81.82
Decision Tree	77.3	77.75	75.86	79.63	85.71
Random Forest	89.36	89.36	89.33	89.39	90.54
Gaussian Naïve Bayes	85.52	85.78	86.49	85.07	86.49

Table 7. Models Evaluations

The Logistic Regression model shows the highest Accuracy (90.07%) and the highest Balanced Accuracy (91%), indicating it is the most reliable for overall performance across all models. It also has the highest Specificity (95.52%) and Precision (95.52%), suggesting it is particularly strong at correctly identifying negative cases and the proportion of positive identifications that are correct. However, its Sensitivity is lower compared to the Support Vector Machine (SVM).

The SVM has the highest Sensitivity (88.73%), which implies it is the best at identifying true positives, but it sacrifices some Specificity and Precision, it has lower overall Accuracy compared to Logistic Regression and Random Forest. The Decision Tree model has the lowest scores in all metrics, indicating that it may be the least effective model among those presented for this dataset.

The Random Forest model presents a good balance across all metrics, with high Accuracy (89.36%), Balanced Accuracy (89.36%), and the second-highest Precision (90.54%). It also has high Sensitivity (89.33%) and Specificity (89.39%), making it a strong contender for the best model. Gaussian Naïve Bayes offers moderate performance across all metrics but does not excel in any metric when compared to Logistic Regression and Random Forest.

In conclusion, the Logistic Regression model appears to be the best model overall due to its superior Accuracy, Balanced Accuracy, Specificity, and Precision.

4. Discussion

Developing a machine learning metamodel for cardiac failure forecasting based on clinical data represents a significant advancement in cardiovascular medicine. This discussion will dive into the key findings of this research, discuss the implications for clinical practice, highlight the strengths and limitations of the metamodel, and suggest avenues for future research.

The evaluation of the proposed metamodel revealed its superior performance compared to other state-of-the-art machine learning models. With an accuracy of 90.07%, the metamodel showcased its potential for accurately predicting heart failure based on clinical test data. This high level of accuracy is promising, as it has the potential to aid healthcare professionals in identifying patients at risk of heart failure and implementing preventive measures in a timely manner. Early detection of heart failure is crucial for initiating appropriate interventions and personalized treatment plans, ultimately leading to improved patient outcomes.

One of the strengths of the metamodel lies in its incorporation of multiple machine learning algorithms, namely Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and Gaussian Naïve Bayes. By blending these algorithms, the metamodel leverages their individual strengths, such as the

ability of decision trees to capture complex interactions and the robustness of Random Forest Classifier in handling noisy data. This integration enhances the metamodel's predictive accuracy and model robustness, making it a valuable tool for forecasting cardiac failure.

Utilizing a combined dataset from five well-known cardiac datasets, including Statlog Heart, Cleveland, Hungarian, Switzerland, and Long Beach, ensures a comprehensive representation of patient characteristics, clinical features, and risk factors. This approach enhances the generalizability and applicability of the metamodel, as it captures a diverse range of patient profiles and healthcare settings. Including 11 standard features from these datasets provides a solid foundation for predicting heart failure, but future studies can explore the integration of additional clinical variables to refine the metamodel's predictive capabilities further.

5. Conclusion

After selecting the best model, there are several potential future works that can be considered, such as:

1. **Feature engineering:** Investigating new features or altering existing ones to possibly enhance the models' predictive abilities. The most pertinent features for heart failure prediction could be found by using feature selection techniques.
2. **Model Optimization:** Improving the performance of the models by fine-tuning the model hyperparameters using methods like grid search or random search. This might entail experimenting with various parameter pairings and assessing how they affect the evaluation metrics.
3. **Ensemble Methods:** Examining ensemble learning strategies to possibly improve the accuracy and robustness of the models, such as combining different models or applying strategies like bagging or boosting.
4. **Handling Unbalanced Data:** Dealing with the problem of unbalanced data, particularly if the heart failure dataset has a sizable class imbalance. To increase the model's capacity to predict heart failure accurately, strategies such as over-sampling the minority class (cases of heart failure) or under sampling the majority class (cases of non-heart failure) could be investigated.
5. **External Validation:** Testing the trained models on different datasets that haven't been seen before to judge how well they generalize. This step is essential to verify the models' performance and dependability beyond the available dataset.
6. **Interpretability and Explainability:** Examining methods for interpreting and explaining the model predictions. To do this and foster confidence in the model, it may be necessary to employ techniques like feature importance analysis or the creation of model-independent explanations.

To sum up, the Logistic regression appears to be the best model for predicting heart failure in this case as it achieves a high score in terms of accuracy, with a value of 90.07%.

Overall, the heart failure prediction dataset offers a useful starting point for developing machine learning models in this domain. However, there are several areas that can be further explored to enhance the predictive performance and practical applicability of the models, including feature engineering to extract more relevant information from the data, model optimization to fine-tune the hyperparameters for improved performance, addressing imbalanced data to prevent biased predictions, external validation to assess the generalizability of the models on unseen data, and enhancing interpretability to gain more insights into the factors contributing to heart failure.

Further research and development efforts in these areas can lead to more accurate and reliable heart failure prediction models, ultimately contributing to enhanced diagnosis and treatment of heart failure patients in real-world scenarios.

References

1. Denolin H., Kuhn H., Krayenbuehl H., Loogen F., Reale A. The definition of heart failure. *Eur. Heart J.* 1983;4:445–448. doi: 10.1093/oxfordjournals.eurheartj.a061500. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
2. Gheorghiade M., De Luca L., Fonarow G.C., Filippatos G., Metra M., Francis G.S. Pathophysiologic targets in the early phase of acute heart failure syndromes. *Am. J. Cardiol.* 2005;96:11–17. doi: 10.1016/j.amjcard.2005.07.016. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
3. Ponikowski P., Anker S.D., AlHabib K.F., Cowie M.R., Force T.L., Hu S., Jaarsma T., Krum H., Rastogi V., Rohde L.E., et al. Heart failure: Preventing disease and death worldwide. *ESC Heart Fail.* 2014;1:4–25. doi: 10.1002/ehf2.12005. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
4. von Haehling S., Anker S.D. Prevalence, incidence and clinical impact of cachexia: Facts and numbers—Update 2014. *J. Cachexia Sarcopenia Muscle.* 2014;5:261–263. doi: 10.1007/s13539-014-0164-8. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
5. Heart A. Heart failure projected to increase dramatically, according to new statistics. *Am. Heart Assoc. News.* 2017;1:2018–2020. [[Google Scholar](#)]
6. Goff D.C., Jr., Brass L., Braun L.T., Croft J.B., Flesch J.D., Fowkes F.G., Hong Y., Howard V., Huston S., Jencks S.F., et al. Essential features of a surveillance system to support the prevention and management of heart disease and stroke: A scientific statement from the American Heart Association Councils on Epidemiology and Prevention, Stroke, and Cardiovascular Nursing and the Interdisciplinary Working Groups on Quality of Care and Outcomes Research and Atherosclerotic Peripheral Vascular Disease. *Circulation.* 2007;115:127–155. [[PubMed](#)] [[Google Scholar](#)]
7. Detrano R., Janosi A., Steinbrunn W., Pfisterer M., Schmid J.J., Sandhu S., Guppy K.H., Lee S., Froelicher V. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am. J. Cardiol.* 1989;64:304–310. doi: 10.1016/0002-9149(89)90524-9. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
8. Ekiz S., Erdoğan P. Comparative study of heart disease classification; Proceedings of the 2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT); Istanbul, Turkey. 20–21 April 2017; New York, NY, USA: IEEE; 2017. pp. 1–4. [[Google Scholar](#)]
9. Nassif A.B., Mahdi O., Nasir Q., Talib M.A., Azzeh M. Machine learning classifications of coronary artery disease; Proceedings of the 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP); Pattaya, Thailand. 15–18 November 2018; New York, NY, USA: IEEE; 2018. pp. 1–6. [[Google Scholar](#)]
10. UCI Machine Learning Repository: Statlog (Heart) Data Set. [(accessed on 1 June 2023)]. Available online: [https://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](https://archive.ics.uci.edu/ml/datasets/statlog+(heart))
11. Bashir S., Khan Z.S., Khan F.H., Anjum A., Bashir K. Improving heart disease prediction using feature selection approaches; Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST); Islamabad, Pakistan. 8–12 January 2019; New York, NY, USA: IEEE; 2019. pp. 619–623. [[Google Scholar](#)]
12. UCI Machine Learning Repository: Heart Disease Data Set. [(accessed on 1 June 2023)]. Available online: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
13. Learning M. Heart disease diagnosis and prediction using machine learning and data mining techniques: A review. *Adv. Comput. Sci. Technol.* 2017;10:2137–2159. [[Google Scholar](#)]

14. Awan S.E., Sohel F., Sanfilippo F.M., Bennamoun M., Dwivedi G. Machine learning in heart failure: Ready for prime time. *Curr. Opin. Cardiol.* 2018;33:190–195. doi: 10.1097/HCO.0000000000000491. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
15. Ahsan M.M., Siddique Z. Machine learning-based heart disease diagnosis: A systematic literature review. *Artif. Intell. Med.* 2022;128:102289. doi: 10.1016/j.artmed.2022.102289. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]