Meta-Learning Based Classification Model for Cardiovascular Disease

Hariharan M (hariharan392@gmail.com)
K. Somasundaram

Method Article

Keywords: Meta Learning, Risk prediction, Machine Learning, SMOTE

Posted Date: April 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2863140/v1

Meta-Learning Based Classification Model for Cardiovascular Disease

Abstract

Cardiovascular disease is a major global health concern and is the leading cause of death and disability worldwide. According to the World Health Organization, cardiovascular disease is responsible for 17.9 million deaths each year, which accounts for 31% of all global deaths. Heart disease is a major cause of mortality worldwide. Machine learning algorithms have shown promise in predicting the risk of heart attacks. Meta-learning is a type of machine-learning method which enables a system to learn how to learn. It involves a set of techniques that allow a system to improve its own learning process. In this paper, we propose a Meta-learning based classification model for Cardiovascular diseases. We consider the dataset (for heart attack classification), which contains 76 attributes with the predicted attribute being the presence of heart disease. We evaluate traditional classification models and Meta-Learning approach for heart attack classification. Additionally, we compared the results using SMOTE and without SMOTE to balance the target classes. The Meta-learning approach outperforms traditional models, providing a more accurate prediction of heart attack risk. These results suggest that the meta-learning approach can be used to improve accuracy.

Keywords: Meta Learning, Risk prediction, Machine Learning, SMOTE.

1. Introduction

Heart attack classification is a significant problem in the medical field, which involves predicting the likelihood of a person experiencing a heart attack based on various risk factors. The problem is important because heart disease is one of the leading causes of death worldwide, and early detection of heart attacks can help save lives and improve patient outcomes. In this paper, we aim to compare the performance of machine learning models with and without SMOTE (Synthetic Minority Oversampling Technique), as well as the performance of meta learning with and without SMOTE, for heart attack classification. We will use a publicly available dataset from UCI repository and evaluate the performance of various machine learning models, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Multi-Layer Perceptron, Voting Classifier. We will also implement a meta learning approach to optimize the performance of these models. The methodology is shown in Fig.1.

1.1 Literature Review

Cardiovascular Diseases (CVDs) are among the leading causes of death and the most common non-communicable diseases worldwide. Anjali et al. [8] developed a real-time detection and warning of cardiovascular disease LAHB for a wearable wireless ECG device. Dwarakana et al. [9] proposed a feature selection with hybrid deep learning based heart disease detection and classification in the e-healthcare environment. In [10], the authors studied the classification of heart murmur using CNN. Bhanu et al. [11] proposed an algorithm for noise reduction of ECG. They extensively used the Chebyshev filter technique and ML based classification model. In [12], a deep learning architecture was proposed for classification of diabetic and normal HRV signals. Zwack et al. [13] given a comprehensive survey on evolution of digital health technologies in this field. Their primary aim is to analyse more than 16,000 articles in this domain based on their scientometric indicators and draw conclusions on these technologies. Gaidai et al. [14] gave different evolution methods of digital health technologies in this filed. Deep learning model for classification of cardiovascular disease using empirical mode decomposition for ECG feature extraction is discussed in [15]. Hamed et al. [16] shown the recent developments in modeling, imaging, and monitoring of cardiovascular diseases using machine learning. An Overview of Deep Learning Methods for Left Ventricle Segmentation was given in [17]. Nahian Ibn et al. [18] proposed a deep learning approach to cardiovascular disease classification employing modified ECG signal from empirical mode decomposition. A meta analysis for effectiveness of artificial intelligence models for cardiovascular disease prediction was given in [19]. Achyut et al. [20] proposed

an ensemble framework for cardiovascular disease prediction. However, these models show some lesser accuracies. This drawback is resolved in this proposed model using meta learning.

THE WORKFLOW

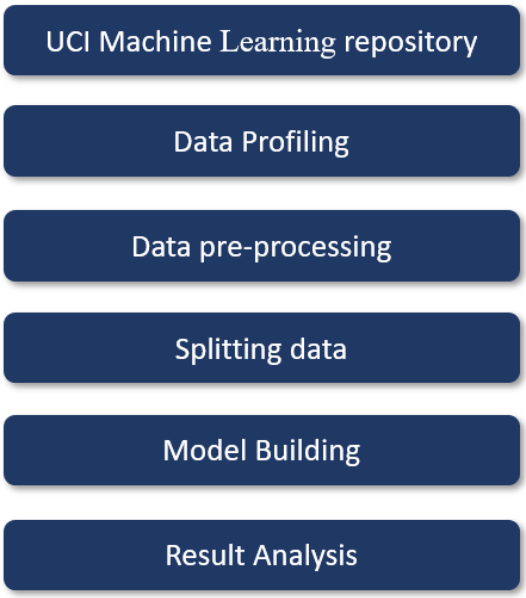

Figure 1: Methodology

2. Data Collection

The publicly available dataset from the UCI repository "Heart Disease" dataset is often used as a benchmark dataset for machine learning algorithms in the field of medical diagnosis. The dataset was originally created in 1988 by researchers at the University of California, Irvine, and it includes four separate databases, each from a different institution: Cleveland, Hungary, Switzerland, and Long Beach V. Each record in the dataset represents a patient who has undergone a variety of medical tests and evaluations, and the goal is to predict whether the patient has heart disease. The predicted attribute is a binary variable, with a value of 0 indicating no heart disease and a value of 1 indicating the presence of heart disease. The dataset contains 76 attributes, including both numerical and categorical variables. However, most published experiments using this dataset focus on a subset of just 14 attributes, which are believed to be the most relevant for predicting

Feature	Description
Age	Age in years
Sex	Sex (1 = male; 0 = female)
Cp	Chest pain type (0 = asymptomatic; 1 = typical angina; 2 = atypical angina; 3 = non-anginal pain)
trestbps	Resting blood pressure (in mm Hg)
Chol	Serum cholesterol in mg/dl
Fbs	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
restecg	Resting electrocardiographic results (0 = hypertrophy; 1 = normal; 2 = having ST-T wave abnormality)
thalach	Maximum heart rate achieved
Exang	Exercise induced angina (1 = yes; 0 = no)
oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment (0 = down sloping; 1 = flat; 2 = upsloping)
Ca	Number of major vessels (0-3) coloured by fluoroscopy
Thal	1 = fixed defect; 2 = normal; 3 = reversable defect
Target	Diagnosis of heart disease (disease status) (Value 0 = Less chances of heart attack; Value 1 = More chances of heart attack)

Table 1: Dataset description

heart disease. The listed features (Table. 1) are age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting ECG, max heart rate, exercise induced angina, old peak, No. of major vessels coloured by fluoroscopy, thallium stress test and target. However, the fact that 70% of the dataset contains duplicates have taken into consideration for cleaning to ensure data quality, and the data is pre-processed to build models and analysis for the data.

3. Data Profiling

Data profiling is an essential step in data preparation and analysis. It involves a range of techniques, including statistical analysis, data visualization, and data mining. It can help to identify missing values, duplicate records, inconsistent data formats, outliers, and other data quality issues that may affect the reliability and usefulness of the data. The dataset is collected

from the UCI repository and imported to perform the basic data profiling for the dataset. The size of the dataset having 1025 records with 13 attributes and a target feature among which 713 are Male and 312 are female. The dataset contains both numerical and categorical features. There are no missing values in the dataset, the visual representation is in Fig. 2. Identifying the unique values of each feature in the dataset is important to understand the range of values each feature can take. This can help in identifying any outliers in the dataset. In the heart attack classification dataset, we found the following unique values for each feature: age (41), sex (2), cp (4), trestbps (49), chol (152), fbs (2), restecg (3), thalach (91), exang (2), oldpeak (40), slope (3), ca (5), thal (4), target (2). The basic statistical description of the dataset for each numerical features, including count, mean, standard deviation, 25%, 50%, 75% percentiles, minimum and maximum values. Duplicate values can impact the accuracy of the model, so it is important to identify and remove them from the dataset. In this dataset, about 723 (70.54%) duplicates have found leaving 302 (29.46%) non-duplicate instances. The duplicates are removed along with the outliers in the dataset in the pre-processing step for further proceedings. The Exploratory Data Analysis (EDA) has done for the entire dataset to get a better understanding of the data.

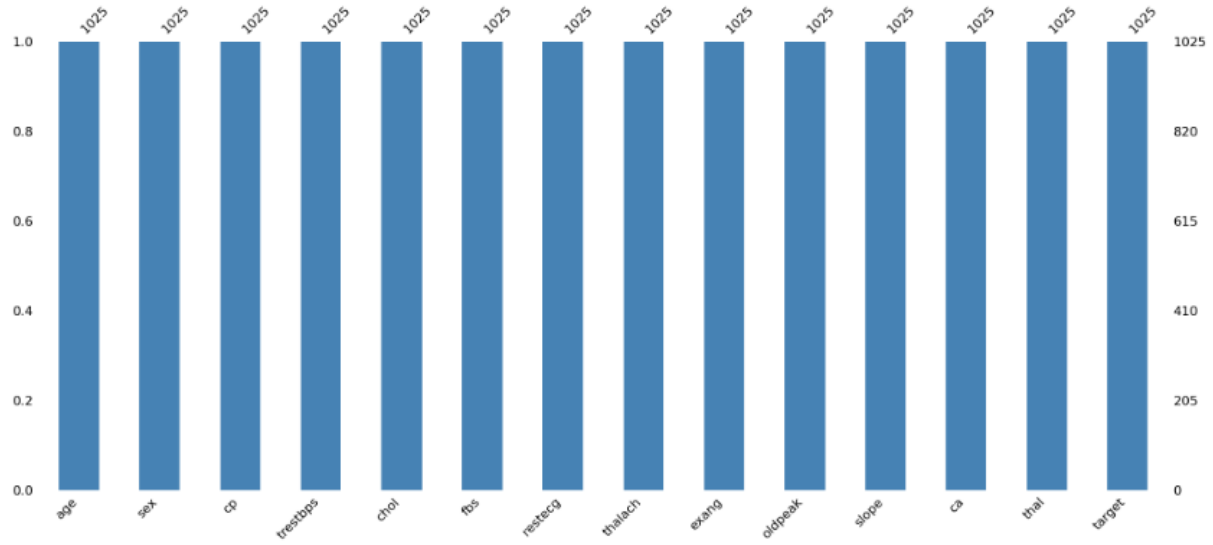

Figure 2: Blank value representation among features.

Now, the features of the dataset are categorized into two types, categorical and numerical. In summary, numerical data is represented by numbers and can be measured or counted, while categorical data is divided into categories or groups and is often represented by words or labels. Categorical data can be further divided into nominal data, which has no inherent order or numerical

value, and ordinal data, which has a specific order or ranking. Categorical features include sex, cp, fbs, restecg, exang, slope, ca, thal, target. Similarly, the numerical features include age, trestbps, chol, thalach, oldpeak.

Categorical features are visualized using count and pie charts shown in figure 3 and 4, while numerical features are visualized using histogram and distribution plots shown in figure 5 and 6. The resulting insights from figures 3 & 4 are for the categorical features, which includes, Sex - 69.6% of the patients are male, 30.4% are female. So, the number of male patients is more than twice that of female patients. Chest Pain Type - Almost half of the patients have an observation value of 0. That is 48.5% is Asymptomatic Angina (Pain without symptoms). On the other half, Atypical angina is in 27.7% of the patients (patients with shortness of breath or non-classical pain). 16.3% of patients have Typical angina (classic exertion pain that comes during any physical activity). Non-anginal pain is the with 7.51%, used to describe chest pain that is not caused by heart disease or a heart attack. Fasting Blood Sugar - Most patients 85.1% have an observation value of 0. That is the Fasting blood sugar of these patients is less than 120 mg/dl. And 14.9% patients have the Fasting blood sugar less than 120 mg/dl. Resting Electrocardiographic - 50% of the patients have Normal resting ECG. 48.5% of the patients have Hypertrophy resting ECG. 1.46% of the patients have ST and T wavelengths. Exercise-induced Angina - Angina is the chest pain caused by the coronary artery's involuntary contraction that feeds the heart. In this context, values 0 are more than twice as values 1. More than half 66.3% of the patients do not have exercise-induced angina. Slope - The minimum observation value is 0 Down sloping with 7.22%. The other two observation values are almost equal. That is 47% have Flat ST wavelength and 45.8% are Upsloping. Ca (number of major vessels (0-3) coloured by fluoroscopy) - 56.4%, the number of large vessels is 0. That is, the number of large vessels coloured by fluoroscopy is absent. After 0 observation value, the other value with the most slices in the pie chart is value 1. The number of large vessels observed in 22% of the patients is 1. Thallium stress test - The thallium stress test is simply an imaging method that evaluates the amount of blood reaching the heart muscle and determines whether a person has coronary artery disease. According to our research, the observation value of 0 is Null. And 53.1% of the patients are Normal, 40% of the patients have Reversible defect, 6.24% of the patients are Fixed defect. Target - 51.3% of the patients have a heart attack risk. 48.7% of the patients have no heart attack risk.

The insights from numerical features in figures 5 & 6 are observed as follows, Age - Most

patients are between 50 and 60. Most patients comes under the age of 58-59. Typically, the general belief is that heart attack increases with age. Resting Blood Pressure - The resting blood pressure of most patients is generally between 110 and 140. There is hefty patient traffic between 120-125, 130-135 and 140-145 values. Cholesterol - Cholesterol value in most patients is between 200-and 280. It peaked with a greater number of patients in the range between 230 and 240. Maximum Heart rate achieved - The maximum heart rate achieved in most patients is between 140-and 175. It peaked with a greater number of patients in the range between 160 and 165. Old peak - Values of most patients in the variable range from 0 to 1.5. Huge set of head counts is under 0-0.1 old peak value. Also interpreted the results for categorical and numerical features with respect to the target class. The summary from the insights, Sex - Female patients are at higher risk for heart attack. Cp - If a patient's "Cp" variable is 1, 2, or 3, we can say that the risk of heart attack is high. Fbs – Both patients with fasting blood sugar lower and greater than 120 mg/dl can be positioned as risky. Restecg - Attention should be paid to patients with a "Rest ECG" value of 1 (Normal). Because the risk of having a heart attack is almost higher than that of not having a heart attack. Exang - If he has exercise-related angina, the probability of not having a heart attack is higher. In contrast, patients who do not have exercise-related angina are more likely to have a heart attack. This means that exercise-related pain has nothing to do with a heart attack. Slope - patients with an observation value of 2 are three times more likely to have a heart attack than not having a heart attack. patients with an observation value of 2 are three times more likely to have a heart attack than not having a heart attack. Ca - the risk of heart attack is almost three times higher in patients with an observation value of 0. Thal - Patients with an observation value of 2 (Normal) are three times more likely to have a heart attack than if they have not. Age - It's just a general thought that older people may be more likely to have a heart attack, but from a graph of the age and target distribution, it's clear that this isn't the case. Resting Blood Pressure - A patient with a " trestbps" value of 130-134 may or may not be likely to have a heart attack. Cholesterol - cholesterol value of 200-250 is precarious for the patients. Maximum Heart rate achieved - The higher the maximum reached heart rate, the higher the probability of the patient having a heart attack. Old peak - the range of 0 to 1.5 is the critical threshold for us. And the correlation between all features and the target variable have resulted, higher levels of chest pain and maximum heart rate achieved are positively correlated with the target variable for heart disease. Conversely, exercise-induced angina and old peak are negatively correlated with the target variable, indicating a lower chance of having a heart attack.

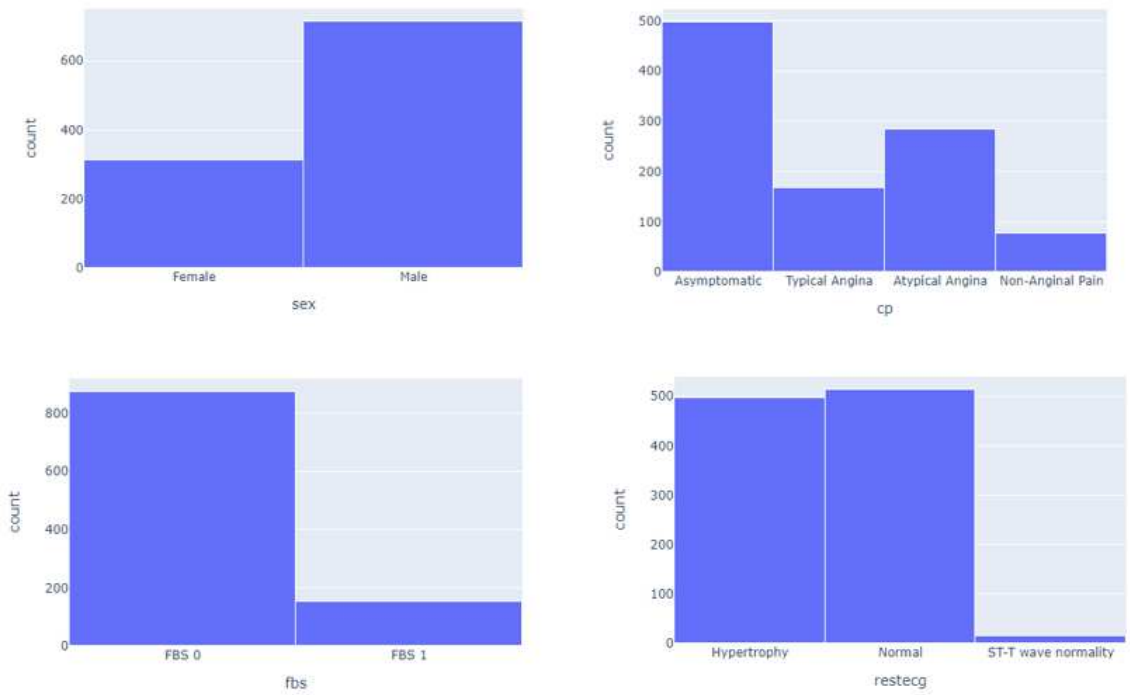

Figure 3(a): Count plot of categorical features

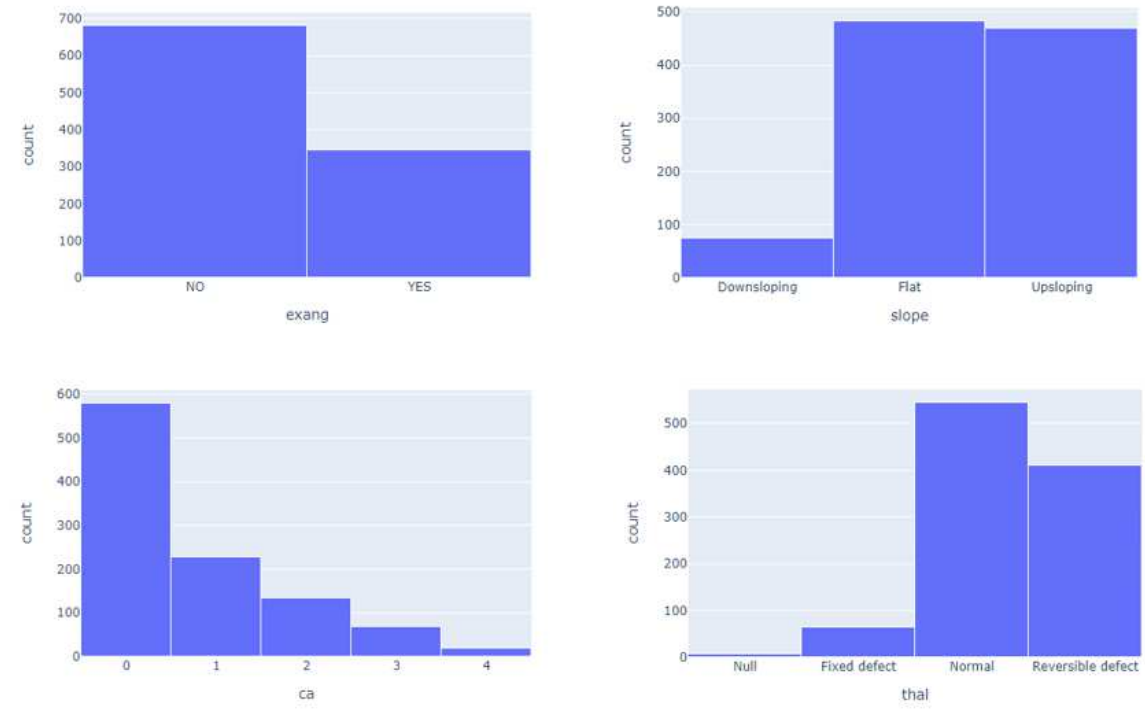

Figure 3(b): Count plot of categorical features

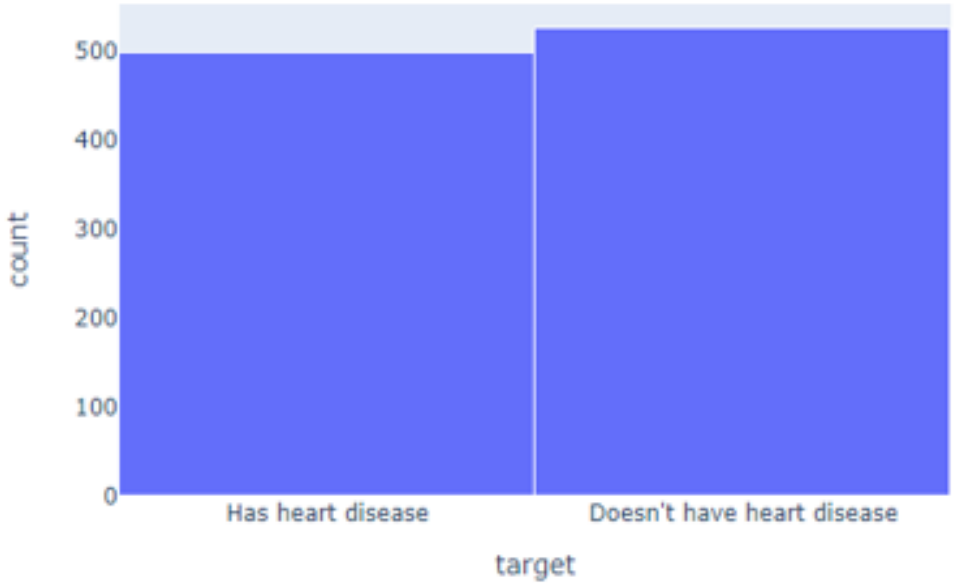

Figure 3(c): Count plot of categorical features

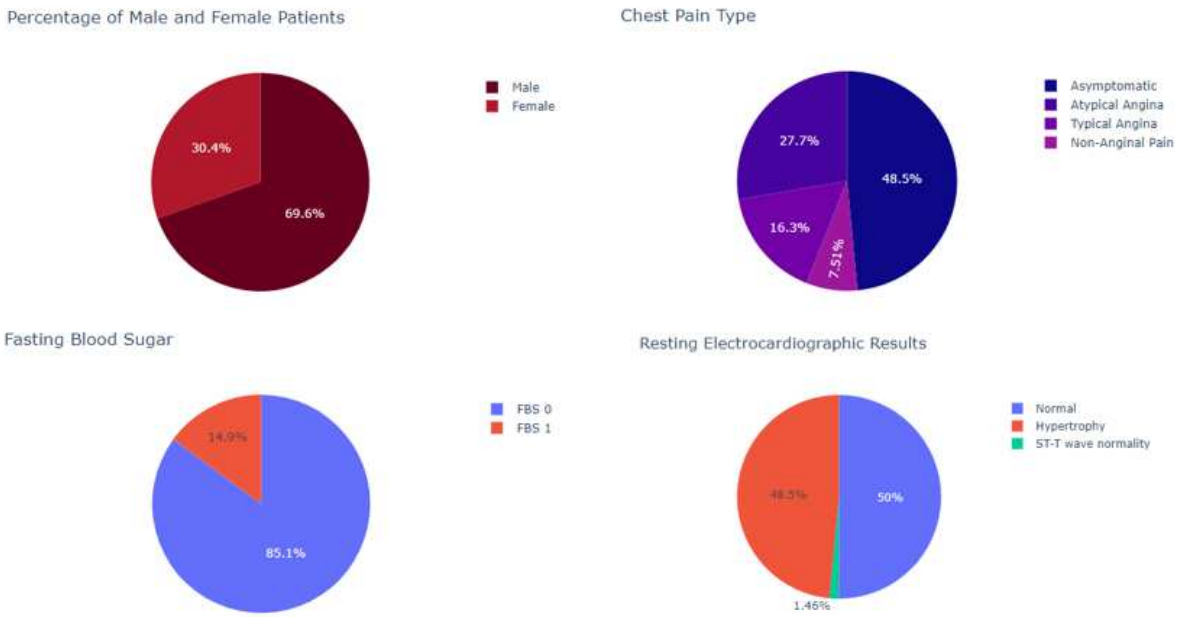

Figure 4 (a): Pie chart for categorical features

Figure 4(b): Pie chart for categorical features

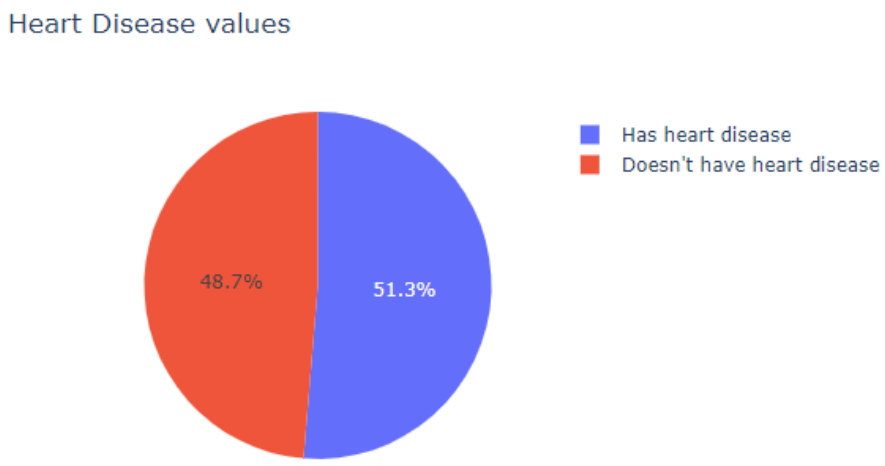

Figure 4 (c): Pie chart for categorical features.

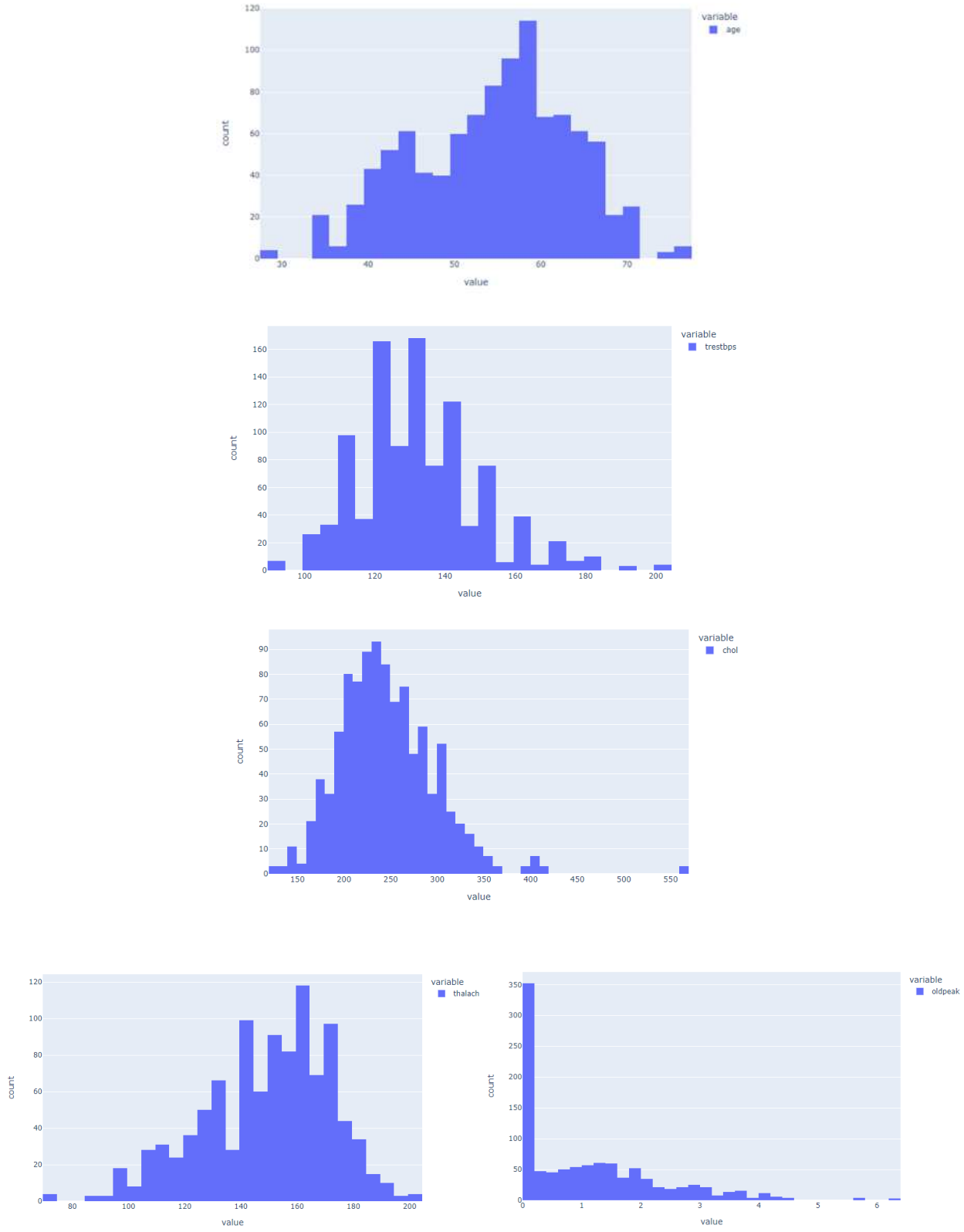

Figure 5: Histogram chart for numerical features.

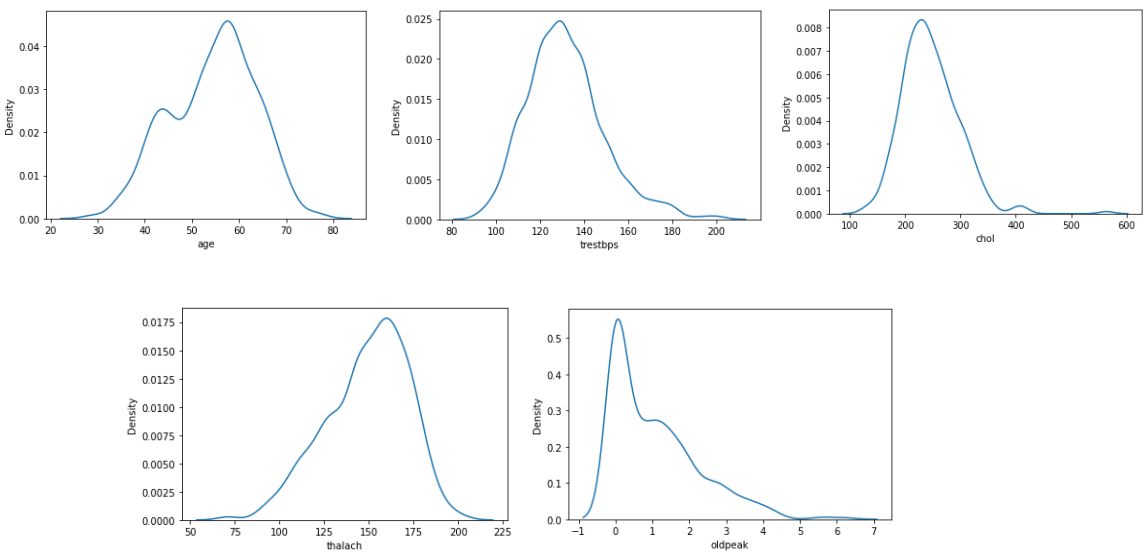

Figure 6: Distribution plot for numerical features.

4. Data Pre-processing

Data pre-processing is a critical step in preparing raw data for analysis or machine learning. It involves a set of techniques to clean, transform, and reformat data to make it suitable for analysis or modeling. Data pre-processing and feature engineering are essential for ensuring the quality and consistency of data, improving the accuracy and effectiveness of analytical models, and avoiding errors and biases in decision-making. The success of any data analysis or machine learning project depends on the quality of data pre-processing. The first step drops large number of duplicate rows in the dataset with the 'inplace' parameter set to 'True', which modifies the original dataset. Next an outlier step is then calculated as 1.5 times the IQR. Any data points that fall outside this range are considered outliers and are removed from the dataset. This step helps to remove any data points that are significantly different from most of the data and may skew the analysis. A 'StandardScaler' object is then created and fit to the specified numeric variables in the dataset. This step scales the numeric variables to have zero mean and unit variance, which is a common step in many machine learning algorithms.

After thorough cleaning, the dataset has been reduced from 1025 records to 283. Out of these, 723 were identified as duplicated records and were removed to ensure the authenticity of the dataset. Additionally, 19 outlier records were also removed. The numerical features in the dataset have been normalized. The cleaned dataset is now ready for model building, albeit with a

smaller sample size.

5. Splitting Data

The first step after pre-processing the data is to split it into a training set (80% of the data) and a testing set (20% of the data) to train and evaluate the machine learning model, respectively. Class imbalance in the dataset are addressed by using SMOTE, which generates synthetic data points for the minority class. K-fold cross-validation can be used to improve the reliability of the model evaluation, with 10 folds being a common choice. This involves splitting the data into 10 equally sized parts, and training and evaluating the model 10 times, with each part being used as the testing set once. Various machine learning algorithms can then be used to train and evaluate the model, with performance evaluated using metrics such as accuracy, precision, recall, and F1-score. Finally, the overall performance of the model is obtained by averaging the evaluation metrics over the 10 folds of the cross-validation procedure.

6. Model Building

In this model building part, the performance of several classification models on a dataset is compared using two different approaches: one with SMOTE oversampling technique and stratified 10-fold cross-validation (CV), and the other without SMOTE and stratified 10-fold CV. The models being compared are: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Multi-Layer Perceptron, Voting Classifier and in Meta Learning. The purpose of comparing these models is to identify the one with the highest accuracy score on this dataset.

In the first approach, SMOTE is applied to the dataset to oversample the minority class (if present) and balance the dataset. Then, the dataset is split into 10 equal parts, and stratified 10-fold cross-validation is performed on each model. Stratification ensures that each fold has the same proportion of target variable classes as the entire dataset. The accuracy score is then calculated for each model. In the second approach, the original dataset is used to train and evaluate the models. The dataset is also split into parts in stratified 80:20 ratio and is performed on each model. The accuracy score is then calculated for each model. By comparing the accuracies of each model using both approaches, we can identify the best-performing model(s) for this specific dataset and problem. It is important to note that the choice of oversampling technique and cross-validation method can affect the accuracy scores of the models. In this case, SMOTE and stratified

10-fold CV were used.

6.1 Logistic Regression

Logistic Regression is a popular classification algorithm in machine learning used for solving binary and multi-class classification problems. It is a statistical model that makes use of the logistic function to predict the probability of a binary outcome based on input features. In logistic regression, the input features are used to calculate a weighted sum, which is passed through the logistic function, also known as the sigmoid function, to produce a probability value between 0 and 1. This probability value represents the likelihood of a given input data point belonging to the positive class. Mathematically, the logistic regression model can be represented as follows:

Let Y be the binary response variable, taking values 0 or 1.

Let X_1, X_2, ..., X_p be the independent variables.

Let β_0, β_1, β_2, ..., β_p be the parameters to be estimated.

The logistic regression model assumes that the log-odds of the response variable being in the category 1 is a linear function of the predictor variables. Mathematically, this can be expressed as:

$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$, where p is the probability of Y being in the category 1. The term logit(p) is the log-odds ratio, which is the logarithm of the ratio of the probability of Y being in category 1 to the probability of Y being in category 0. This log-odds ratio is transformed to the probability scale using the logistic function: $p = \frac{1}{1+e^{-\text{logit}(p)}}$. The logistic function "squashes" the log-odds ratio to the range [0, 1], which represents the probability of Y being in category 1.

6.2 Decision Tree

Decision Tree is a simple and powerful classification algorithm that can be used for both binary and multi-class classification problems. It can handle both numerical and categorical data and can be easily visualized and interpreted. However, it may be prone to overfitting and can be sensitive to small changes in the input data, which can be addressed through techniques such as pruning, ensemble methods, and regularization. Entropy is nothing but the uncertainty in our dataset or measure of disorder. The formula for Entropy is,

$Entropy = \sum_{i=1}^{c} - \mathrm{p_i} * \log_2(p_i)$, here c is number of classes.

$Gini = 1 - \sum_{i=1}^{c}(p_i^2)$, information gain measures the reduction of uncertainty given some feature and it is also a deciding factor for which attribute should be selected as a decision node or root node. Information Gain = $E(Y) - E(Y/X)$.

6.3 Random Forest

Random Forest is a popular machine learning algorithm used for solving classification problems. It is an ensemble learning method that combines multiple decision trees to make a final prediction. Random Forest works by building multiple decision trees, each one trained on a randomly selected subset of the input features and a random subset of the training data. Each decision tree is built using the same process as the decision tree algorithm, selecting the best split at each node based on a criterion such as entropy or Gini impurity. Once all the decision trees are built, they are used to make a prediction for each new input data point. Each tree's prediction is counted as a vote, and the final prediction is based on the majority vote of all the decision trees. This voting process helps to reduce the risk of overfitting and improves the accuracy of the model. Random Forest has several advantages over a single decision tree, such as reducing overfitting, handling noisy data, and improving the accuracy of the model. It is a powerful and versatile algorithm that can be used for both binary and multi-class classification problems. Overall, Random Forest is a robust and reliable classification algorithm that combines the strengths of multiple decision trees to produce a more accurate prediction.

6.4 Support Vector Machine

A support vector machine (SVM) is a type of machine learning algorithm used for classification and regression analysis. SVM works by finding the best possible decision boundary (called the hyperplane) that separates the data into different classes. The hyperplane is chosen in such a way that it maximizes the margin between the two classes, which makes the SVM more robust to noisy data. In other words, SVM tries to find the best line or plane that separates the data points into different classes by maximizing the distance between the line and the nearest data points of each class. It finds the optimal hyperplane by maximizing the margin distance between the observations of the classes using the Hinge loss function. The dimension of the hyperplane depends upon the number of input features. If the number of features is *N*, then the dimension for the hyperplane is *N*-1. Here, *x* is the input variable, *t* is the target variable and *w* are the model parameters.

$$l(y) = \max(0, 1 + \max_{y \neq t} \boldsymbol{w_y X} - \boldsymbol{w_t X})$$

6.5 Naïve Bayes

The Naive Bayes algorithm is a type of machine learning algorithm used for classification tasks. It is based on the Bayes theorem, which states that the probability of a hypothesis (in this case, a class label) is proportional to the probability of the evidence (in this case, a set of features or attributes) given the hypothesis.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c)\, P(x_2|c) \ldots . P(x_n|c)P(c)$$

Where, P(c | x) is the posterior probability,P(x | c) is the likelihood,P(c)is the class prior probability, $P(x)$ is the predictor prior probability.

The "naive" part of the algorithm comes from the assumption that all features are independent of each other, which makes the calculations much simpler. This assumption may not always be true in real-world scenarios, but in many cases, it still produces accurate results. Naive Bayes works by calculating the probability of each class label given the set of features, and then choosing the class label with the highest probability as the predicted label for a new instance. The algorithm can handle both categorical and numerical data, making it useful for a wide range of applications.

6.6 Multi-Layer Perceptron

A multi-layer perceptron (MLP) is a type of neural network that consists of multiple layers of interconnected nodes (called neurons) arranged in a feedforward manner. It is a type of supervised learning algorithm that can be used for both classification and regression tasks. Each neuron in an MLP takes input from the previous layer, applies a non-linear activation function, and outputs a signal to the next layer. The layers in between the input and output layers are called hidden layers, and they can learn complex representations of the input data through a process called backpropagation. The row vectors of the inputs and weights are $x = [x_1, x_2, \ldots, x_n]$ and $w = [w_1, w_2, \ldots, w_n]$ respectively. Hence, the summation is equal to the dot product of the vectors x and w,

$$z = x.w + b$$

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

To know an estimation of how far we are from our desired solution, a loss function is used.

$$MSE_i = (y_i - \hat{y}_i)^2$$

Loss function is calculated for the entire training dataset and their average is called the *Cost function* **C**.

$$C = MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

MLP is a flexible and powerful algorithm that can handle non-linear relationships between the input and output variables. The performance of an MLP depends on the number of hidden layers, the number of neurons in each layer, the choice of activation function, and the learning rate.

6.7 Voting Classifier

A voting classifier is an ensemble learning algorithm that combines the predictions of multiple individual classifiers to make a final prediction. In a voting classifier, each individual classifier is trained on the same training set but uses a different algorithm or a different set of hyperparameters. When making a prediction, the voting classifier collects the predictions of all individual classifiers and then chooses the class label that has the most votes. There are two main types of voting classifiers: hard voting and soft voting. In hard voting, each individual classifier casts a single vote for its predicted class label, and the class label with the most votes is chosen as the final prediction. In soft voting, each individual classifier assigns a probability score to each class label, and the final prediction is the class label with the highest average probability score across all individual classifiers. Voting classifiers can be used for both classification and regression tasks, and they can improve the accuracy and robustness of the predictions compared to individual classifiers. They are especially useful when the individual classifiers have different strengths and weaknesses and can complement each other's performance.

6.8 Meta Learning

Meta learning, also known as "learning to learn," is a type of machine learning that focuses on developing algorithms that can learn how to learn. In other words, meta learning is about creating

models that can learn from previous learning experiences and use that knowledge to adapt to new tasks more quickly and efficiently. Figure 7 shows the proposed model approach for meta learning.

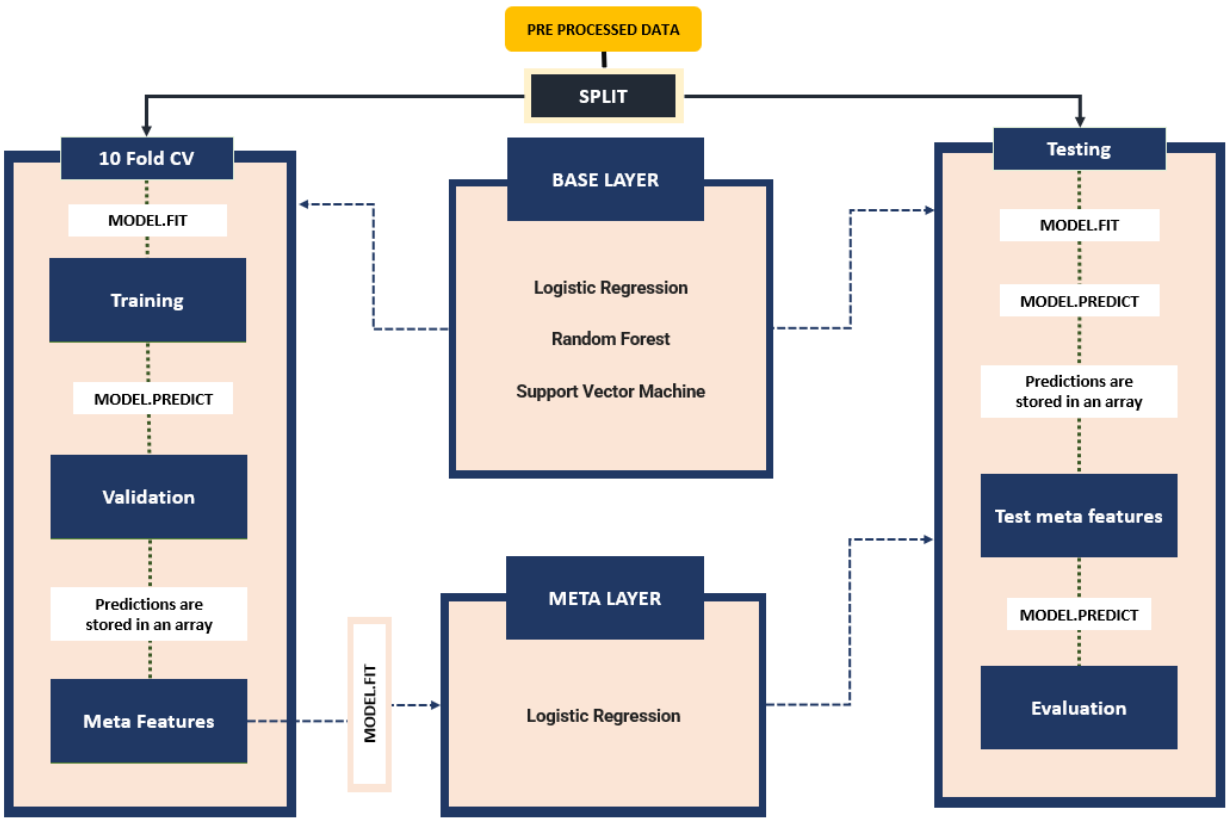

Figure 7: Flow diagram of the proposed model.

Meta-learning is a subfield of machine learning that focuses on learning how to learn, i.e., using past experience to improve the ability of a model to learn new tasks. Meta-learning algorithms typically learn from a set of tasks, called the training tasks, and then apply the learned knowledge to new tasks, called the test tasks. It then defines three base models, i.e., logistic regression, random forest, and support vector machine, as well as a meta-model, which is another logistic regression model. The number of folds for cross-validation is set to 10, and an array is initialized to store the meta-features, which are the predictions of the base models for each fold of the cross-validation. A stratified k-fold cross-validation is defined to ensure that each fold has a similar distribution of classes. The base models are trained, and meta-features are generated in a nested loop. For each

base model, the following steps are performed for each fold: The data is split into training and validation sets. The model is fit on the training data. The model predicts on the validation data. The predictions are stored in the meta-features array. Once all base models are trained and meta-features are generated, the meta model is trained on the meta-features. A new array is initialized to store the meta-features for the test data (test_meta_features). This array has a number of rows equal to the length of the test data and a number of columns equal to the number of base models. The base models are then trained on the entire input data (x and y) and predictions are generated for the test data. These predictions are stored in the test_meta_features array. Finally, the meta model predicts on the test meta-features and the accuracy of the predictions is calculated.

In summary, meta-learning is an approach where multiple models are stacked, trained on data to generate features, and then a meta-model is trained on those features to make a final prediction. The main idea is to learn from the performance of the base models and use that information to improve the final prediction.

7. Results and Analysis

After building machine learning models, their performance needs to be evaluated. Figure 8 and Table 2 illustrates a comparison between the stratified 10-fold cross-validation scores of two types of models: one that uses Synthetic Minority Over-sampling Technique (SMOTE) and the other that does not use SMOTE. The blue bars represent the cross-validation scores of the model that uses SMOTE, while the red bars represent the cross-validation scores of the model that does not use SMOTE. Stratified 10-fold cross-validation involves dividing the data into 10 equal parts, training the model on 9 parts, and testing it on the remaining 1 part. This process is repeated 10 times with each part used as the test set once. The results are then averaged to obtain a more accurate estimate of the model's performance. The use of SMOTE is a technique used to address class imbalance in the dataset, where one class is significantly smaller than the other. It involves generating synthetic samples of the minority class to balance the dataset. The comparison in Figure 8 and Table 2 shows how the performance of the two models differs when SMOTE is used or not. The blue bars demonstrate that the model that uses SMOTE has a higher cross-validation score compared to the red bars, which means that it performs better.

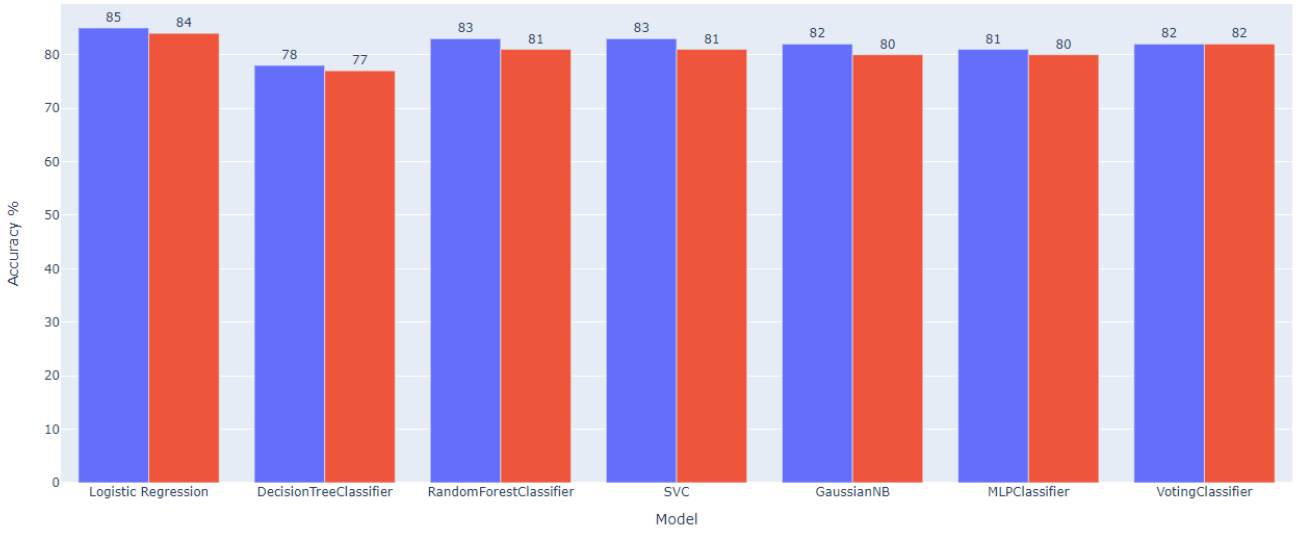

Figure 8: Comparison of Cross-validation scores.

	Model	CV Score	CV
0	Logistic Regression	85	84
1	DecisionTreeClassifier	78	77
2	RandomForestClassifier	83	81
3	SVC	83	81
4	GaussianNB	82	80
5	MLPClassifier	81	80
6	VotingClassifier	82	82

Table 2: Comparison of Cross-validation scores.

After preprocessing the data and splitting it into stratified training and testing sets, models are built, and their performance is evaluated. In this process, stratified 10-fold cross-validation (cv) is not used. Fig.9 and Table 3 displays a comparison between the accuracies of two sets of models. One set is built using a technique called SMOTE (Synthetic Minority Over-sampling Technique) and is represented by the blue bars in the figure. The other set is built without using SMOTE and is represented by the red bars. SMOTE is a technique used to address class imbalance in machine learning datasets. Class imbalance occurs when one class (usually the minority class) is

underrepresented in comparison to the other classes. This can lead to biased models that perform poorly on the minority class. SMOTE works by creating synthetic examples of the minority class, which helps to balance the dataset and improve the performance of the models.

The results in Figure 9 and Table 3 indicate that the models built using SMOTE (blue bars) almost have higher or equal accuracies compared to the models built without SMOTE (red bars). This suggests that the use of SMOTE has improved the performance of the models by addressing the class imbalance issue in the dataset. It should be noted that in this process, stratified 10-fold cv was not used to evaluate the performance of the models.

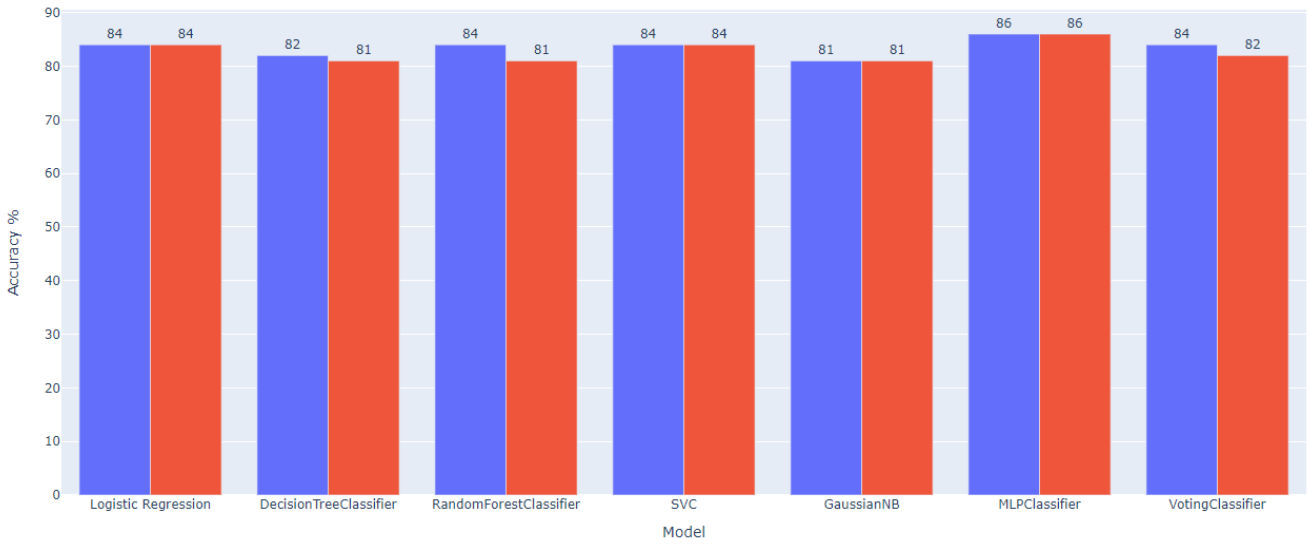

Figure 9: Comparison of Accuracy across the models.

	Model	Acc	Accuracy
0	Logistic Regression	84	84
1	DecisionTreeClassifier	82	81
2	RandomForestClassifier	84	81
3	SVC	84	84
4	GaussianNB	81	81
5	MLPClassifier	86	86
6	VotingClassifier	84	82

Table 3: Comparison of Accuracy across the models.

The results of the stratified 10-fold cross-validation scores for the meta learning model are compared in Figure 10 and Table 4, with and without the use of SMOTE. The purpose of this comparison is to determine the effect of SMOTE on the model's performance. The meta learning approach have yielded better performance results from the traditional methods. The meta learning model itself gives a better result when the target class is treated using SMOTE to balance the dataset.

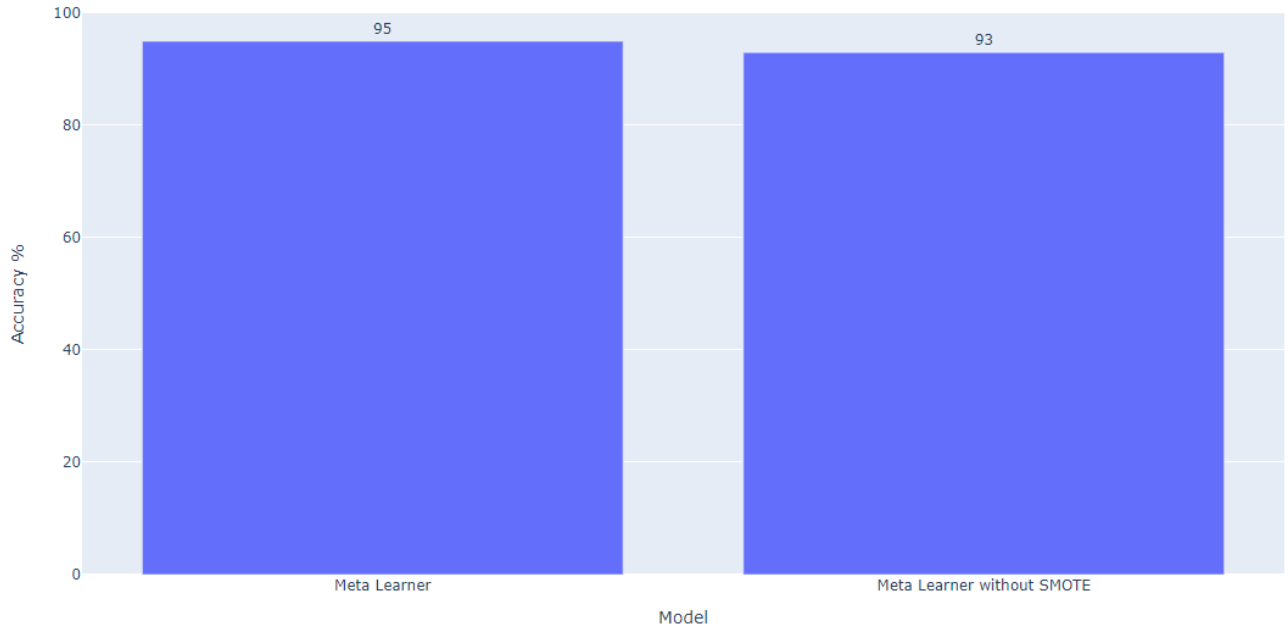

Figure 10: Comparison of Accuracy across the models.

	Model	Accuracy
0	Meta Learner	95
1	Meta Learner without SMOTE	93

Table 4: Comparison of Accuracy across the models.

Conclusion

Meta-learning is a modern and effective approach to heart attack classification that can improve the accuracy and efficiency of diagnosis. The ability of meta-learning algorithms to learn from multiple datasets and adapt to new data with minimal retraining makes it a promising tool for healthcare professionals. With the increasing prevalence of heart disease worldwide, incorporating meta-learning techniques in medical research and practice can help to reduce the burden of heart attacks on individuals and healthcare systems. However, further research is needed to optimize the use of meta-learning in clinical settings and to ensure its safety and efficacy. Overall, meta-learning presents a valuable opportunity to improve heart attack classification and ultimately, to enhance patient outcomes. In summary, meta-learning presents a valuable opportunity to enhance patient outcomes and improve heart attack classification. We hope that this research will inspire further studies in this area to continue exploring the potential of meta-learning in healthcare.

Conflicts of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] A. Saran Kumar, R. Rekha, An improved hawks optimizer based learning algorithms for cardiovascular disease prediction, Biomedical Signal Processing and Control, Volume 81, 2023, 104442, ISSN 1746-8094.

[2] Shaw, S.K., Patidar, S. (2023). Heart Disease Diagnosis Using Machine Learning Classification Techniques. In: Ranganathan, G., Fernando, X., Rocha, Á. (eds) Inventive Communication and Computational Technologies. Lecture Notes in Networks and Systems, vol 383. Springer, Singapore. https://doi.org/10.1007/978-981-19-4960-9_35

[3] C. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-6, doi:

10.1109/ASET53988.2022.9734880.

[4] Katarya, R., Meena, S.K. Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. Health Technol. 11, 87–97 (2021)

[5] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020).

[6] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[7] Patel, Jaymin & Tejalupadhyay, Samir & Patel, Samir. (2016). Heart Disease prediction using Machine learning and Data Mining Technique. 10.090592/IJCSC.2016.018.

[8] Arunan, Anjali, Rahul Krishnan Pathinarupothi, and Maneesha Vinodini Ramesh. "A real-time detection and warning of cardiovascular disease LAHB for a wearable wireless ECG device." In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 98-101. IEEE, 2016.

[9] Kallimani, Jagadish S., Ranjan Walia, and Birhanu Belete. "A Novel Feature Selection with Hybrid Deep Learning Based Heart Disease Detection and Classification in the e-Healthcare Environment." *Computational Intelligence and Neuroscience* 2022 (2022).

[10] Sudarsanan, Sukanya, and J. Aravinth. "Classification of heart murmur using CNN." In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 818-822. IEEE, 2020.

[11] Prakash, M. Bhanu, V. Sowmya, E. A. Gopalakrishnan, and K. P. Soman. "Noise Reduction of ECG using Chebyshev filter and Classification using Machine Learning Algorithms." In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 434-441. IEEE, 2021.

[12] Swapna, G., R. Vinayakumar, and K. P. Soman. "Diabetes detection using deep learning algorithms." *ICT express* 4, no. 4 (2018): 243-246.

[13] Zwack, Clara C., Milad Haghani, Matthew Hollings, Ling Zhang, Sarah Gauci, Robyn Gallagher, and Julie Redfern. "The evolution of digital health technologies in cardiovascular disease research." *npj Digital Medicine* 6, no. 1 (2023): 1.

[14] Gaidai, Oleg, Yu Cao, and Stas Loginov. "Global cardiovascular diseases death rate prediction." *Current Problems in Cardiology* (2023): 101622.

[15] Li, Ya, Jing-hao Luo, Qing-yun Dai, Jason K. Eshraghian, Bingo Wing-Kuen Ling, Ci-yan Zheng, and Xiao-li Wang. "A deep learning approach to cardiovascular disease classification using empirical mode decomposition for ECG feature extraction." *Biomedical Signal Processing and Control* 79 (2023): 104188.

[16] Moradi, Hamed, Akram Al-Hourani, Gianmarco Concilia, Farnaz Khoshmanesh, Farhad R. Nezami, Scott Needham, Sara Baratchi, and Khashayar Khoshmanesh. "Recent developments in modeling, imaging, and monitoring of cardiovascular diseases using machine learning." *Biophysical Reviews* (2023): 1-15.

[17] Shoaib, Muhammad Ali, Joon Huang Chuah, Raza Ali, Khairunnisa Hasikin, Azira Khalil, Yan Chai Hum, Yee Kai Tee, Samiappan Dhanalakshmi, and Khin Wee Lai. "An Overview of Deep Learning Methods for Left Ventricle Segmentation." *Computational intelligence and neuroscience* 2023 (2023).

[18] Hasan, Nahian Ibn, and Arnab Bhattacharjee. "Deep learning approach to cardiovascular disease classification employing modified ECG signal from empirical mode decomposition." *Biomedical signal processing and control* 52 (2019): 128-140.

[19] Baashar, Yahia, Gamal Alkawsi, Hitham Alhussian, Luiz Fernando Capretz, Ayed Alwadain, Ammar Ahmed Alkahtani, and Malek Almomani. "Effectiveness of artificial intelligence models for cardiovascular disease prediction: network meta-analysis." *Computational intelligence and neuroscience* 2022 (2022).

[20] Tiwari, Achyut, Aryan Chugh, and Aman Sharma. "Ensemble framework for cardiovascular disease prediction." *Computers in Biology and Medicine* 146 (2022): 105624.