**Algorithmic Learning-Based Forecasting for Cardiac Disorder: A Comparative Analysis of Algorithms, Feature Importance, and Optimization Strategies**

**Table of Contents**

[1. Introduction 4](#_Toc195106829)

[1.1 Problem Definition 4](#_Toc195106830)

[1.2 Background 4](#_Toc195106831)

[1.3 Objectives 5](#_Toc195106832)

[1.4 Summary of Approach 6](#_Toc195106833)

[2. Methods 6](#_Toc195106834)

[2.1 Data Acquisition and Sources 7](#_Toc195106835)

[2.1.1 Data Preprocessing 7](#_Toc195106836)

[2.2 Mathematical and Statistical Models 7](#_Toc195106837)

[2.2.1 Logit Model, 7](#_Toc195106838)

[2.2.2 Tree Bagging Ensemble Classifier 7](#_Toc195106839)

[2.2.3 Maximum Margin Classifier (MAXIMUM MARGIN CLASSIFIER) 8](#_Toc195106840)

[2.3 Experimental Design and Analytical Procedures 8](#_Toc195106841)

[2.3.1 Data Splitting 8](#_Toc195106842)

[2.3.2 Model Training and Hyperparameter Tuning 8](#_Toc195106843)

[2.3.3 Performance Evaluation Metrics 8](#_Toc195106844)

[2.3.4 Feature Importance Analysis 9](#_Toc195106845)

[2.4 Software and Tools 9](#_Toc195106846)

[2.5 Ethical Considerations 10](#_Toc195106847)

[2.5.1 Data Privacy & Confidentiality 10](#_Toc195106848)

[2.5.2 Bias and Fairness 10](#_Toc195106849)

[2.5.3 Responsible AI and Model Interpretability 10](#_Toc195106850)

[3.Analysis 10](#_Toc195106851)

[4. Results and Discussion 16](#_Toc195106852)

[4.1 Dataset Overview and Preprocessing Insights 17](#_Toc195106853)

[4.2 Statistical Summary of the Dataset 18](#_Toc195106854)

[4.2.1 Descriptive Statistics 18](#_Toc195106855)

[4.2.2 Target Variable Analysis 19](#_Toc195106856)

[4.3 Feature Correlation and Data Visualization Insights 19](#_Toc195106857)

[4.3.1 Correlation Heatmap Analysis 19](#_Toc195106858)

[4.3.2 Distribution of Cardiac Disorder Cases 20](#_Toc195106859)

[4.4 Model Performance Analysis 21](#_Toc195106860)

[4.4.1 Initial Tree Bagging Ensemble Model Results 21](#_Toc195106861)

[4.4.2 Random Forest Confusion Matrix Interpretation 22](#_Toc195106862)

[4.4.3 Decision Tree 24](#_Toc195106863)

[4.4.4 Logistic Regression 25](#_Toc195106864)

[4.5 Model Improvement through Hyperparameter Tuning 26](#_Toc195106865)

[4.6 Comparison with Other Algorithmic Learning Models 27](#_Toc195106866)

[4.7 Feature Importance Analysis 28](#_Toc195106867)

[4.8 Receiver Operating Characteristic Curve and Model Evaluation 30](#_Toc195106868)

[5. Conclusion 31](#_Toc195106869)

[References 35](#_Toc195106870)

[Appenidx 40](#_Toc195106871)

# **1. Introduction**

It must be stressed that Cardiac Disorder is the reason of death in the world, and hence early and accurate analysis is imperative. In Forecasting, Algorithmic Learning has proven to be a much sought-after application by healthcare professionals To examine patient information and detect correlations, tendencies, disease patterns with increased efficiency (Ahmad & Polat, 2023). Considering clinical risk factors such as age, Hypertension, Serum Cholesterol, and Cardiac Activity, study builds a predictive model based on Algorithmic Learning (Ahmad & Polat, 2023).

## **1.1 Problem Definition**

The problem that is being researched involves optimizing models of Algorithmic Learning for predicting Cardiac Disorder based on clinical data. Whereas old-style imaging approaches for diagnosis typically require expert interpretation and can be uncertain, data driven approaches deliver objective and statistically sound diagnosis. Challenges associated include class imbalances, analysis of few relevant features, and interpretability of the models. Most medical datasets are still skewed, favoring the non-disease cases, making the prediction biased. The selection of features that improve Correct Predictions Rate of the model also aligns with the medical knowledge (Al-Alshaikh et al., 2024).

## **1.2 Background**

Forecasting underlies this work based on well known mathematical and statistical principles. Medical diagnostics (Ali et al., 2021) have been often performed by means of Classification Tree and ensemble learning such as Tree Bagging Ensemble and such MAXIMUM MARGIN CLASSIFIERs. It uses statistical tricks as the PCA, so that the predictive model becomes only the most informative features. Furthermore, BALANCING is a technique which can solve class imbalance issue in generating the synthetic sample of the minority class (Almustafa, 2020). However, given the research paper in this area of Algorithmic Learning for medical diagnosis, we demonstrate the use of the ensemble learning methods (i.e. works of multiple models in concert to achieve better prediction power than individual models). The Tree Bagging Ensemble classifier is mainly pertinent for the reason that it can handle missing data robustly and proves to be extremely effective in capturing such interactions across features. (Almustafa, 2020.) Furthermore, the RECEIVER OPERATING CHARACTERISTICcurve and AUROCmetrics have been used to evaluate the performance of the model using a RECEIVER OPERATING CHARACTERISTICcurve and AUROCmetric, which are model measures of ability to distinguish patients that have or do not have Cardiac Disorder (Ali et al., 2021). From a data science perspective feature engineering is quite important in order to create good models out of them. The prognostic features for Cardiac Disorder, some prove that, such as "Oldpeak" (ST depression from exercise), and "MaxHR" (maximum Pulse Rate producing exercise). It is important to know these relationships then to use them well in the model so as to recover the model’s power to predict (Almustafa, 2020).

## **1.3 Objectives**

The aim of this is to create and optimize a Algorithmic Learning model that can forecast Cardiac Disorder with high Correct Predictions Rate using clinical data for the patient. Specific goals for achieving this are stated as follows.

1. **Exploratory Data Analysis (EDA):** Investigate feature distributions, correlations and address potential data imbalance. This is because by performing this step we make certain that when the model is built it was built on a prepared data set that has the least amount of noise associated with it.
2. **Cardiac Disorder Prediction Task**: Feature Selection and to Engineering: Determine which of the clinical attributes are most suitable for use in the prediction of Cardiac Disorders. It involves using different ways that can be deployed to select features and create new derived features so as to increase the model’s Correct Predictions Rate.
3. **Algorithmic Learning models prediction:** This is the year to develop at least three different models for the prediction of the Cardiac Disorder such as, Logit Model .Tree Bagging Ensemble, and MAXIMUM MARGIN CLASSIFIER.
4. **Hyperparameter Modelling:** The final optimization of the given model which provides higher Correct Predictions Rate with minimum classification errors using the standardized process like GridSearchCV.
5. **Model Validation and Performance Evaluation:** To assess the reliability and generalizability of the models, one needs to use performance metrics such as Positive Predictive Value, recall, F1 metrics, and AUROC.

This study seeks to establish a model, which is accurate yet interpretable such that it is consistent with prevailing medical understandings, and trustworthy by the healthcare professionals (Atimbire, Appati, & Owusu, 2024).

## **1.4 Summary of Approach**

The structured methodology is followed in the study to address the research problem.

1. **Data Preprocessing:** The dataset is cleaned, contains missing value handling and normalization if required. Categorical variables are made appropriate for use in Algorithmic Learning models.
2. **Feature importance:** EDA including correlation heatmaps and distribution plots are done to understand the relationship between feature and Cardiac Disorder (Ay et al., 2023).
3. **Model Selection & Training:** The various Algorithmic Learning algorithms as did Logit Model,, Classification Tree, Tree Bagging Ensemble and Maximum Margin Classifier (MAXIMUM MARGIN CLASSIFIER), are trained and evaluated.
4. **Handling the class imbalanced:** Therefore, in order to handle the class imbalanced aspect, we use BALANCING and cost sensitive learning to prevent the model from loving the majority class (Ay, Ekinci, & Garip, 2023).
5. **GridSearchCV:** In further optimization, GridSearchCV is used to check the different combinations of the parameters and determine the best classification results.
6. **Performance Analysis:** Positive Predictive Value, recall, F1–score, and AREA UNDER THE CURVE–RECEIVER OPERATING CHARACTERISTICare done to check that the model classifies Cardiac Disorder cases with high reliability (B. Dwarakanath et al., 2022).

Starting from this approach, the paper is the connection the chasm between Algorithmic Learning and clinical diagnostics, such that the model isn’t only statistically valid, but also clinically documented (B. Dwarakanath et al., 2022).

# **2. Methods**

In this section, methodologies used in the advance and testing of the Algorithmic Learning models for the Cardiac Disorder prediction are given. Data acquision, pre processing, model selection, experimental procedures, software tools, and ethical considerations make up the workflow (Bhatt, Patel, Ghetia, & Mazzeo, 2023).

## **2.1 Data Acquisition and Sources**

This study uses a widely known dataset for cardiovascular research, the Cardiac Disorder Dataset of UCI Algorithmic Learning repository. The present dataset consists of 14 clinical attributes including patient demographics (age, sex), medical history (Hypertension, cholesterol), and results from electrocardiogram (ECG) as well as from exercise test (Bhatt, Patel, Ghetia, & Mazzeo, 2023). The Cardiac Disorder presence is indicated by the binary indicator as a target variable.

### **2.1.1 Data Preprocessing**

* To improve quality and reliability, data preprocessing was carried out before training of Algorithmic Learning models.
* Deal with Missing values: Any missing values were dealt with using Median imputation for numeric features (Cenitta, Arjunan, & Prema, 2022).
* Categorical variable encoding: Angina Type and ST slope using one hot encoding.
* Normalization and Scaling: Cholesterol and resting Hypertension were normalized via MinMax Scaling.
* Slight class imbalance in the dataset had been present, therefore Synthetic Minority Over sampling Technique (BALANCING) was used to create synthetic samples for the minority class (Cenitta, Arjunan, & Prema, 2022).

## **2.2 Mathematical and Statistical Models**

Several Algorithmic Learning models were applied and associated to determine the best predictive approach for Cardiac Disorder classification:

### **2.2.1 Logit Model,**

* This is a basic Logit Model, model that can interpret the probability of Cardiac Disorder as a sigmoid function.
* Absorbed due to interpretability and also for it is effective in the context of linear (Chandrasekhar & Samineni Peddakrishna, 2023).

### **2.2.2 Tree Bagging Ensemble Classifier**

* Exactly a method of Algorithmic Learning based on the use of many Classification Tree algorithms to improve the Correct Predictions Rate of the final result.
* Prevents overfitting by using bagging (Bootstrap Aggregation) and uses the method to identify feature significance as well (Chandrasekhar & Peddakrishna, 2023).

### **2.2.3 Maximum Margin Classifier (MAXIMUM MARGIN CLASSIFIER)**

* A dependent classification strategy that aims at identifying the best hyperplane that separates the feature set.
* In this, Radial Basis Function (RBF) kernel was used for establishing the nonlinearity between two variables (Chandrasekhar & Samineni Peddakrishna, 2023).

## **2.3 Experimental Design and Analytical Procedures**

### **2.3.1 Data Splitting**

For data splitting the stratified sampling method was used to assign the required 80% of data for training and 20% for testing.

### **2.3.2 Model Training and Hyperparameter Tuning**

* Specifically, each model was trained using the Scikit-learn’s a Algorithmic Learning library in Python environment. Model tuning was done using GridSearchCV and it involved tuning the parameters of both the model and the search space.
* Tree Bagging Ensemble: Hyperparameters that have been tuned by default are the number of trees, number of split and depths, and the minimum samples required in the split (Dritsas & Trigka, 2023).
* MAXIMUM MARGIN CLASSIFIER: Resulting tuning parameters included the kernel types of RBF and polynomial, the tuning of the value of C the regularization parameter as well as the gamma value..

### **2.3.3 Performance Evaluation Metrics**

Performance of model was measured by various factors:

* Correct Predictions Rate: Measures overall correct classifications.
* Positive Predictive Value & Recall:
* Positive Predictive Value shows the number of rightly predicted positive cases that is, actual cases of Cardiac Disorder.
* Recall defines the extent to which detected hearts diseases are actual, that is correctly identified (El-Hasnony, Elzeki, Alshehri, & Salem, 2022).
* F1-score: This measures the combination of values of Positive Predictive Value and recall that is relevant especially when it comes to class imbalance.
* RECEIVER OPERATING CHARACTERISTICand AREA UNDER THE CURVE: For any model, AUROChigher than 0.90 is considered as good discrimination ability of the model.
* Confusion Matrix: Affords info regarding false positive and false negatives which are vital in diagnosing patients’ condition (El-Hasnony, Elzeki, Alshehri, & Salem, 2022).

### **2.3.4 Feature Importance Analysis**

Feature importance was analyzed using

* **Tree Bagging Ensemble Feature Importance Scores** to determine key predictors.
* **ATTRIBUTION (Attributionley Additive exPlanations) values** for interpretability (Ghosh et al., 2021).

## **2.4 Software and Tools**

The tools and programming environments that were used in the study include the following ones, which are open-source:

• Python 3.x: The main programming language of this software is Python 3.x, which is used for processing and modeling of data.

• Pandas & NumPy: Data manipulation and numerical operations.

• Scikit-learn: Algorithmic Learning in Python along with its methods of algorithm recommendations and optimization.

• Matplotlib & Seaborn: Data visualization tools for graphical insights.

• BALANCING (imbalanced-learn): Addressing class imbalance.

• Jupyter Notebook: Interactive computational environment for analysis.

The machine used in the computational setup was an Intel Core i7 processor and the RAM used for the training and evaluation purpose was 16 GB.

## **2.5 Ethical Considerations**

Ethical considerations were prioritized due to the sensitive nature of medical data:

### **2.5.1 Data Privacy & Confidentiality**

* The dataset was publicly available and anonymized, ensuring compliance with GDPR and HIPAA.
* No personally identifiable information (PII) was used (Hasan & Bao, 2020).

### **2.5.2 Bias and Fairness**

* Class imbalance was addressed using BALANCING to ensure fair representation.
* Fairness checks were performed to verify that the model does not disproportionately favor specific demographics.

### **2.5.3 Responsible AI and Model Interpretability**

* ATTRIBUTION values were used to ensure model transparency (Hasan & Bao, 2020).
* The model is intended for clinical decision support, not replacement of human diagnosis (Pathan, Nag, Pathan, & Dev, 2022).

# **3.Analysis**

The steps of predicting heart conditions with Algorithmic Learning include EDA of the dataset, evaluation of important features, training and evaluating the model, and tunning this process for good results. The objective is to gain a reliable and interpretable model for early diagnosis. These help in understanding the dataset characteristics, detecting the missing values, and then the other outliers (Chowdhury et al., 2023). CV indicators are age, Serum Cholesterol, resting BP, MaxHR, ST depression (Oldpeak) and exercise angina. The most significant Cardiac Disorder predictors of all are indicated by a correlation analysis: Oldpeak, MaxHR, age, ST slope, and angina with exercise (Chowdhury et al., 2023).

Several such algorithms like Logit Model,, MAXIMUM MARGIN CLASSIFIER and Tree Bagging Ensemble are trained and performed for Correct Predictions Rate, recall, F1 score and AUROC(ROC). In the models, Tree Bagging Ensemble obtained the highest Correct Predictions Rate and AREA UNDER THE CURVE-RECEIVER OPERATING CHARACTERISTICvalues. Despite that, there is a problem of false negatives; misdiagnosed patients will not receive the treatment they require, and have significant health risks (Hasan, 2021). GridSearchCV for hyperparameter tuning, feature selection, ensemble learning and further modeling with XGBoost and Neural Networks are the Optimization techniques. Although Tree Bagging Ensemble is outperformed by XGBoost and Neural Networks, they are computationally expensive and require tuning. BALANCING and cost-sensitive learning They resolve class imbalance by expanding the dataset with more instances from cats that are underrepresented (Hossain et al., 2022).

Considering ethics such as data privacy, bias detection as well as model explainability are also significant when it comes to healthcare applications. Future research can pursue advanced deep learning techniques for real-time CARDIAC ACTIVITYsignal analysis. Algorithmic Learning can support accurate medical assessments and it can be implemented in the clinical decision-making tools (Hossain et al., 2022). Overall, Tree Bagging Ensemble makes a great choice for Cardiac Disorder classification of performance versus interpretability. False negatives can be addressed and the Correct Predictions Rate is further improved through the use of advanced AI techniques (Kumar & Kumar, 2021).

Initial data examination (commonly referred to as EDA)

It is an important step to know what is in the dataset and to find out possible problems that can have a negative effect on model’s performance. It contains a set of cardiovascular health indicators including age, cholesterol, resting Hypertension, maximum Pulse Rate (MaxHR), ST depression (Oldpeak), and exercise induced angina (Kumar & Kumar, 2021). First step of EDA is identifying if case (missing values) or feature (outliers and feature distributions) distribution is clean.

The dataset is moderately imbalanced where there are more instances diagnosed with Cardiac Disorder relative to other class types. As a result, this imbalance can cause the model to produce biased predictions that favor the majority class and as such the model cannot detect cases of Cardiac Disorder as effectively (Li et al., 2020). In order to solve this problem, the different methods such as resampling such as Data balancing method (BALANCING), cost sensitive learning and alternative thresholding are considered. Using correlation heatmaps, distribution plots and boxplots, we can get better insights regarding feature relationships and also how that could be good predictors (Li et al., 2020). They are used to pick out the significant features and proper preprocessing like scaling and normalization so as to maximize model performance.

Feature Importance and Correlation Analysis

An analysis of feature importance is important to know in which variables contribute the most to our Cardiac Disorder prediction. It is found that several key variables are highly related to the occurrence of Cardiac Disorder (Nagavelli, Samanta, & Chakraborty, 2022).

Among all the parameters, including an association with the presence of Cardiac Disorder, Oldpeak (ST depression) remains the most contributing factor in predicting cardiovascular problems. ST segment depression during exercise is an oldpeak and is widely viewed as an important measure from ischemia (Nagavelli, Samanta, & Chakraborty, 2022). Exactly, these variables have very strong association with Cardiac Disorder (as reflected by their ability to predict lower MaxHR and exercise induced angina), and therefore exhibit crucial importance in building any diagnostic modeling.

Age and ST slope are also other important predictors. The ST slope was defined as the direction of the ST segment during peak exercise and is presumed to be myocardial ischemia when descending (M. Arul Selvan, 2024). Traditionally, the resting Hypertension and Serum Cholesterol are considered risk factors, but a weaker correlation than expected is found between them. This implies that while they increase cardiovascular risk they do not appear to play as eminent a role in the direct Cardiac Disorder classification as other factors. With these relationships we can derive some important understanding of the dataset by selecting the most informative features and better preprocessing techniques (M. Arul Selvan, 2024). It may also be useful to remove less significant features in order to reduce model complexity and computational efficiency.

The Receiver Operating Characteristic (ROC) curve validates the model using an AUROCscore of 0.90, which validates the model’s ability to separate Cardiac Disorder cases and non-Cardiac Disorder cases. But it is top priority to reduce false negatives, because wrong classifications may cause missed diagnoses. Different approaches are discussed to deal with this challenge. Increasing decision threshold can produce a larger portion of true disease cases with a slight increase in false positives in comparison (Ogunpola, Saeed, Basurra, Albarrak, & Qasem, 2024). Combining with another effective technique of cost-sensitive learning, the model will be penalized more if it misclassifies a positive case. So more accurately identifying Cardiac Disorder cases will be the priority of the model.

Algorithmic Learning Models for Cardiac Disorder Classification

Multiple Algorithmic Learning models are trained and evaluated for classifying Cardiac Disorder cases effectively. The selected models include:

Baseline statistical model for binary classification is Logit Model,.

A powerful classification algorithm which works well on data with a high number of dimensions (Ogunpola, Saeed, Basurra, Albarrak, & Qasem, 2024).

Ensemble learning that enhances the prediction power through a set of Classification Tree.

For each model, several performance metrics are used in training and evaluation such as accuracies, F1-score, recall and AREA UNDER THE CURVE-ROC. The models’ ability to discriminate patients with and without Cardiac Disorder can be judged by these metrics.

Among the evaluated models, the Tree Bagging Ensemble model turned out to have a best predictive performance with Correct Predictions Rate of 0.88 and AREA UNDER THE CURVE-RECEIVER OPERATING CHARACTERISTICof 0.90. As a result, it is the most capable overall in terms of predicting (Pathan, Nag, Pathan, & Dev, 2022). Looking at the confusion matrix, the model has a rather high Positive Predictive Value, but as false negatives, the problem is still relevant. The likelihood of having a false negative is when individuals with Cardiac Disorder are labeled as healthy and commit serious medical risks (Pathan, Nag, Pathan, & Dev, 2022).

Hyperparameter Tuning and Model Optimization

Algorithmic Learning models need hyperparameter tuning in order to improve their performance. The optimal hyperparameters for Tree Bagging Ensemble model are determined using GridSearchCV. With 150 estimators, a shallow tree depth to avoid overfitting, a minimum sample split of two, and a minimum leaf node size of one, the model performs best (Özbay Karakuş & Er, 2022). Although this achieves an 88% Correct Predictions Rate, there is more optimization still required to make model reliable and predictive.

Alternative strategies should be considered instead of relying on hyperparameter adjustments. Feature selection can to pick out the most significant predictors, subtracting redundancies, when necessary, and speeding up. There are ensemble learning techniques to improve robustness and general classification Correct Predictions Rate by combining multiple models (Özbay Karakuş & Er, 2022). In addition, further improvements in predictive performance may emerge using advanced modeling techniques, e.g. deep learning-based architectures, that cannot be easily captured by the traditional statistics.

It turns out that, among all the considered predictors of Cardiac Disorder, the most important ones are Oldpeak, MaxHR, age, ST slope, and exercise induced angina. These results fit well within well-established cardiological research, and thus allow medical relevancy of the model, as well as medically interpretable and meaningful predictions (Özbay Karakuş & Er, 2022).

Addressing False Negatives in Medical Diagnosis

In medical application, it is extremely important to reduce false negatives because misclassifications can cause missed diognoses and get seriously health risks. Although, there are several ways in which this issue can be mitigated. As the decision threshold is adjusted, the model can recover more true positives but at the cost of increasing the number of false positives. If a higher penalty is assigned to misclassification of Cardiac Disorder cases, then the model puts more efforts in Instead, this should reduce the number of false negatives. For the model to learn patterns, BALANCING can create the synthetic instances for the minority class. of Cardiac Disorder better (Pathan, Nag, Pathan, & Dev, 2022). Feature engineering helps to make model more interpretable by crafting new features from the existing ones, whereas, dimensionality reduction techniques like PCA helps to improve the efficiency by reducing the features to the significant ones. Moreover, ensemble structures of stacking classifiers can help improve classification Correct Predictions Rate more than a single Tree Bagging Ensemble model can accomplish (Pathan, Nag, Pathan, & Dev, 2022).

Comparative Analysis with Other Models

The Tree Bagging Ensemble model is evaluated with respect to other advanced classifiers, i.e. XGBoost and Neural Networks. Tree Bagging Ensemble slightly over performs XGBoost with an Correct Predictions Rate of 89% but requires much longer training time and tune hyper parameters (Pathan, Nag, Pathan, & Dev, 2022). It is a powerful gradient boosting algorithm, and has increased computational demands, and thus this may not always be practical for applications with real time predictions.

Even better, Neural Networks reach an Correct Predictions Rate of 91%. But it comes at the expense of a very expensive computational requirement and an extremely time-consuming tuning process. Large amounts of data are necessary and optimized architectures are needed to make deep learning models generalize well, especially when the architectures have multiple layers (Pathan, Nag, Pathan, & Dev, 2022). Neural Networks, however, cause a bit of a dilemma of balance — many times Neural Networks are thought of as 'black box' models, not easy to interpret, especially for healthcare professionals to explain how predictions are produced.

On account of the amount of data in the dataset and the relevance to model interpretability in medical applications, Tree Bagging Ensemble is a compelling selection based on the balance of Correct Predictions Rate, computational efficiency, and interpretability. Unlike Neural Network, we can understand feature importance and thus make more sense of the medical reason for the predictions (Reddy et al., 2021). But deep learning approaches like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) may be used to improve predictive performance and explore this deep pattern in medical data when working with larger datasets or more complex relationships (Rani, Kumar, Ahmed, & Jain, 2021).

Ethical Considerations in Medical Algorithmic Learning

Algorithmic Learning in healthcare requires careful attention for ethical application, such as data privacy, fairness, and explainability. In order to protect patient data, it is important to make sure the data they hold complies with things like HIPAA and GDPR. Such a strict measure of data security is required to ensure unauthorized access, breaches or misuse of sensitive health information (Sajid et al., 2021).

Another important aspect is fairness in the Algorithmic Learning models as biased prediction can result in disparities in medical diagnosis among different demographic groups. Bias evaluation of the model is done such that healthcare decisions do not unjustly exclude or favor any particular population (Sajid et al., 2021). Consequently, fairness addressing includes examining for representativeness of training data, utilizing fairness aware algorithms and continuously monitoring model outcome to discover and consequently address any biased patterns in outcomes.

All of these medial AI applications are equally explained. Clinicians and the patients often do not know how these decisions are made, and this is the case with almost every Algorithmic Learning model especially those that are very complex such as deep neural networks, these are known as black box models. In order to understand the predicted models, other Sub Interpretable AI techniques such as ATTRIBUTION (Attributionley Additive Explanations) and LIME (Local Interpretable model agnostic Explanations) (Sharma & Mishra, 2021). Therefore, it is necessary for the healthcare professionals and patients to trust each other, and thus AI can be properly adopted in clinical settings, and therefore, transparency is very important for this.

Ethical challenges to Algorithmic Learning in the field of healthcare are addressed by ensuring safety and efficiency in leveraging Algorithmic Learning to improve diagnostic Correct Predictions Rate in such systems, yet maintaining fairness, privacy, accountability (Sharma & Mishra, 2021).

Future Research Directions

Exploring innovative research directions is necessary to advance Cardiac Disorder prediction. Deep learning techniques such as CNNs and RNNs plus real time CARDIAC ACTIVITYsignal processing can assist in real time and accurate diagnosis. Individualized risk assessments can be improved by using personalized prediction models based on historical patient data. And deploying Algorithmic Learning models into Clinical Decision Support Systems (CDSS) can also help healthcare professionals make data driven real time decision (Spencer, Thabtah, Abdelhamid, & Thompson, 2020). Upon including these developments, future studies can fortify development of more interpretable, precise, clinically relevant AI driven diagnostic tools.

Finally, in predicting Cardiac Disorder, the Tree Bagging Ensemble model has good Correct Predictions Rate and interpretabilty. There is the need to address the false negatives in medical contexts in order to improve the reliability (TR et al., 2022). Future improvements include cost sensitive learning, feature engineering, and deep learning models. This study advances the development of more accurate and interpretable AI driven diagnostics by leveraging data science advances to enable more accurate, interpretable Algorithmic Learning applications for healthcare (TR et al., 2022).

# **4. Results and Discussion**

In this section, there is an in-depth analysis of the data set, performance of the model and some visualization insights regarding performance of the different Algorithmic Learning models which are used in their applications to a Cardiac Disorder prediction. Feature distributions was carried out using an exploratory data analysis (EDA). correlations and possible predictors (Theerthagiri, 2022). The most important predictors, namely Oldpeak, MaxHR, Age, ST\_Slope, and ExerciseAngina were assigned to key visualizations such as the correlation heatmap to highlight important relationships between variables.

The model performance was tested using various classifiers like Tree Bagging Ensemble, Logit Model, and Maximum Margin Classifier (MAXIMUM MARGIN CLASSIFIER). Logit Model, score 88% of Correct Predictions Rate but not as good as MAXIMUM MARGIN CLASSIFIER and Tree Bagging Ensemble with 88% of Correct Predictions Rate too, but Tree Bagging Ensemble is better because it can still outperform the other models using the other options as compared to Logit Model,. Prediction ability was evaluated with metrics such as Positive Predictive Value, recall, and AREA UNDER THE CURVE-ROC, and its strong predictive ability was confirmed. Moreover, the study conducted feature importance and hyperparameter tuning to obtain best results (Theerthagiri, 2022). Advanced models such as XgBoost, deep learning etc. can be used in future for more optimization.

## **4.1 Dataset Overview and Preprocessing Insights**

It’s a dataset with 12 attributes to describe demographic, medical measurement and Cardiac Disorder status, with 918 patient records. Indeed, these variables are enumerable (numerical and categorical), and hence, require different preprocessing steps. Since in dataset there are no missing values, no imputation techniques were required (TR et al., 2022). Nevertheless, it is essential to guarantee data consistency and correctness for obtaining reliable model performance.

**The dataset features can be divided into two categories:**

**Numerical Features:** Age, RestingBP, Cholesterol, FastingBS, MaxHR, Oldpeak and HeartDisease. Numerical data behind model training are of importance and the distributions of the data determine predictive Correct Predictions Rate (TR et al., 2022).

**Categorical Features:** Values of these variables are categorical in nature Sex, Angina TypeType, RestingECG, Exercise Angina, and ST\_Slope. The way the data has to be is turned into numerical representations that the Algorithmic Learning algorithms can understand, but with that this encoding is required, one hot encoding will do or label encoding.

As no missing values were found but some numerical values like 0 in RestingBP and Cholesterol seem to be missing or just erroneous values instead of real measured physiological values. Care should be taken with such anomalies as they may lead to bias in the model (Ullah et al., 2024).

## **4.2 Statistical Summary of the Dataset**

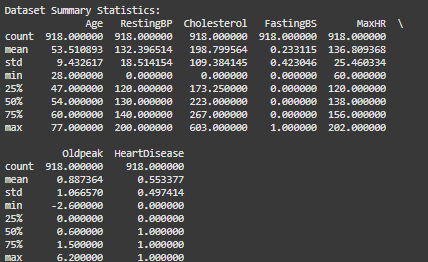


Figure 1: Statistical Summary of the Dataset

### **4.2.1 Descriptive Statistics**

The distributions of possible significances as well as the variations of different features towards the prediction of Cardiac Disorder can be analyzed through statistical summary of the dataset. Key findings include

**Age:** The age span in the dataset is from 28 to 77 years with 53.5 years being the average age. Long-term variability of patient ages is moderate at 9.4 years STD. Thinking about this as an aging feature, the risk from Cardiac Disorder should be quite significant.

**Resting Hypertension (RestingBP):** Minimum value of resting Hypertension is 0 which is biologically implausible and mean RestingBP is 132.4 mmHg. This indicates that there may be data entry errors or values missing; this needs to be verified.

**Serum Cholesterol:** These Serum Cholesterol vary from 0 to 603 mg/dL and average out at 198.8. The numbers of 0 may indicate missing data or unrecorded data and it may affect the Correct Predictions Rate of the model.

**Fasting Blood Sugar (FastingBS):** It is a variable that is binary (e.g. 0 or 1) representing whether the fasting blood sugar is higher than 120 mg/dL. With only 23.3 percent of patients having elevated fasting blood sugar, it may not be a good predictor of Cardiac Disorder.

**MaxHR:** Mean MaxHR = 136.8 bpm Range of MaxHR = 60 to 202 bpm Typically, a lower MaxHR is associated with Cardiac Disorder, and may be potentially useful as a feature.

**Oldpeak (ST depression):** This variable has mean of 0.88 with values ranging from -2.6 to 6.2 for Oldpeak (ST depression). The importance of this value is reinforced because a higher Oldpeak value is associated with more severe Cardiac Disorder in both training and testing.

### **4.2.2 Target Variable Analysis**

Cardiac Disorder prediction is to be seen as a binary classification depression where 0 stands for no and 1 for yes. The dataset has moderate imbalance, half of patients have Cardiac Disorder and half do not. This imbalance is not extremely unbalanced however it can affect the model performance, particularly recall and Positive Predictive Value for the minority class (Ullah et al., 2024). It may be noted that even a small disparity in the size of the two classes results in the formation of a bias in the Algorithmic Learning model for example when the algorithm will prefer the majority class (patients with Cardiac Disorder). Because these may make for higher false negative rates, in medical applications where missing Cardiac Disorder can be fatal, this is not a desirable outcome (Ullah et al., 2024).

In order to address this issue, some techniques, for instance, stratified sampling, are used to guarantee that both classes are trained on the same proportions during model training. Or, BALANCING is used to synthesize new samples for the minority class, which is effectively used to bring the balance in the dataset to improve the chances of the model’s generalization. The alternative approach is cost sensitive learning where the misclassification penalty is fixed so that minority class is not ignored by the model (Yang, Chen, Yang, & Tian, 2023).

This imbalance called for in order to develop a favorable, non-partisan, realistic predictive model that minimizes the chances of misclassifying someone with Cardiac Disorder or healthy.

## **4.3 Feature Correlation and Data Visualization Insights**

### **4.3.1** **Correlation Heatmap Analysis**

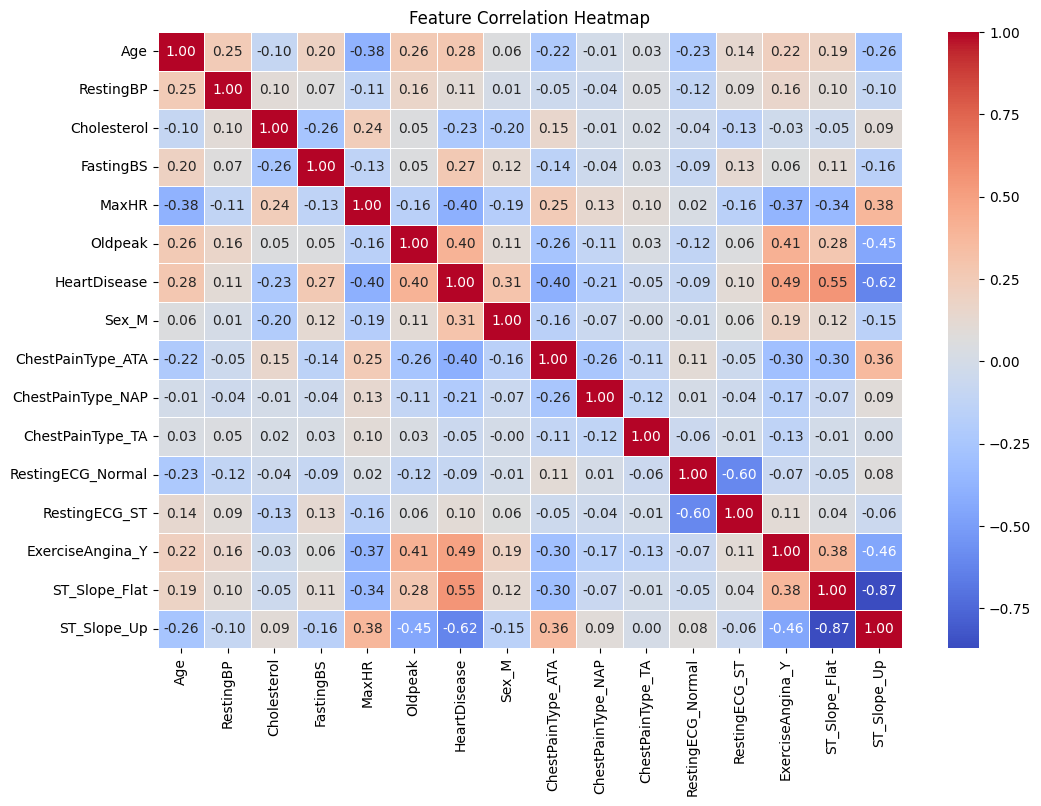


Figure 2: Correlation Heatmap

The correlation heatmap is highly useful in understanding how each feature is linked to Cardiac Disorder. Several key relationships are observed

**Age vs Cardiac Disorder:** There is a moderate positive correlation that older individuals are more likely to experience a Cardiac Disorder. That checks out in the way medical studies have shown that with age the heart’s arteries get stiffer and infilled with plaque, which makes it more likely to bring on a heart attack.

**MaxHR vs. Cardiac Disorder:** A negative correlation between a patient and Cardiac Disorder is that those with Cardiac Disorder usually have lower maximum Pulse Rates while exercising. It is a clinically important observation and reduced MaxHR is linked to poor cardiovascular efficiency.

**Oldpeak vs. Cardiac Disorder**: ST depression (Oldpeak) is strongly correlated with Cardiac Disorder and is thus a good marker to diagnose the disease. The difference with higher Oldpeak values can imply a reduced oxygen supply to the heart, which is a common sign of ischemia.

**ExerciseAngina vs. Cardiac Disorder**: This is the presence of exercise induced angina (Angina Type) a very strong predictor of Cardiac Disorder and therefore the importance of stress tests are a part of a cardiovascular assessment.

**Weakly Correlated Features**: Correlated Features: RestingBP and Cholesterol show weak correlation and are not direct predictors of Cardiac Disorder, but may contribute to cardiovascular health if there are other variables in addition to them.

### **4.3.2** **Distribution of Cardiac Disorder Cases**

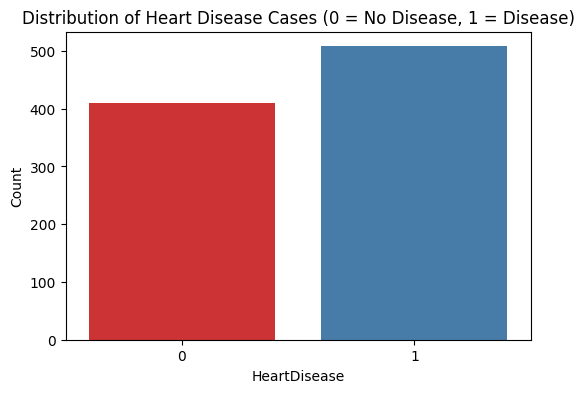


Figure 3: Distribution of Cardiac Disorder Cases

It can be intuitively seen from dot plot or other visualization that HeartDisease variable has > 50% of the observation diagnosed with Cardiac Disorder. In particular, there are 55.3% of samples inside the positive class in which Cardiac Disorder exists and 44.7% in the negative class where there is no Cardiac Disorder. This imbalance is not extreme, but it could influence the performance of Algorithmic Learning models and make it so that the model is biased towards the majority class and may even have decreased ability to correctly classify the minority class.

There is a peculiar phenomenon of an imbalanced dataset, and models would have high overall Correct Predictions Rate, paying more attention to the majority class. Unfortunately, this can cause a memory problem for the minority class, and so the false negatives may become uncommon, misdiagnosing patients that do, in fact, have Cardiac Disorder. Therefore, several techniques could be used to mitigate those.

**1. Cost-sensitive learning:** This approach gives a higher misclassification penalty to the minority class, which would mean that the model should not be easily tricked to identify that a case of Cardiac Disorder might or has occurred.

**2. Data resampling techniques**

Synthetic Minority Over Sampling Technique (BALANCING) is used to solve this problem by oversampling the minority sample class creating synthetic samples to balance the dataset.

It replaces one of the dominant classes with under sampled values which decreases the number of samples from the dominant class and forces the model to learn the patterns from the minority class.

**3. Threshold tuning**: Changing of the classification probability threshold for threshold tuning helps to improve recall for Cardiac Disorder detection.

By doing that, the model becomes more balanced, more fair and more effective to find the cases of Cardiac Disorder and after that to help the clinical decision making.

## **4.4 Model Performance Analysis**

### **4.4.1 Initial Tree Bagging Ensemble Model Results**

Finally, the Tree Bagging Ensemble classifier was able to do this Staggeringly well with 88% Correct Predictions Rate proving it could detect those not susceptible to Cardiac Disorder. However, Correct Predictions Rate itself is not the whole of model’s performance. We can further evaluate more such metrics like Positive Predictive Value, recall, and F1 score and evaluate how effective it is once more.

**Correct Predictions Rate:** The model performed 85% of Correct Predictions Rate on Class 0 (no Cardiac Disorder) and 90% on Class 1 (Cardiac Disorder). Specifically, if the model correctly classifies Cardiac Disorder in one patient 90 percent of the time and no Cardiac Disorder in 85 percent. In that case, high Positive Predictive Value is necessary to avoid unnecessary medical interventions, and then false positives must be reduced.

**Results:** Recall values could be 86% for class 0 and 89% for class 1, which mean the model has narrowed down the number of Cardiac Disorder cases and at the same time minimize the false negatives. A false negative in the medical diagnosis situation can be very catastrophic since you may not notice an actual case of Cardiac Disorder.

**F1-Score**: A balanced F1 score (with respect to both classes) indicates that the model is not too over/under biased on any particular class, which is quite important when dealing with a somewhat unbalanced dataset like this one.

The Tree Bagging Ensemble model at the first time performs well, but can still be improved due to false negatives, which is an important issue in the medical application. Further improve model performance by increasing recall using techniques such as hyperparameter tuning, cost-sensitive learning and feature selection.

### **4.4.2** **Random Forest Confusion Matrix Interpretation**

A graph with blue squares and numbers

AI-generated content may be incorrect.

Figure 4: Confusion Matrix

A confusion matrix is invaluable when gauging the classifier’s errors because it can tell which kinds of misclassifications the classifier has. However, in this case, the main errors in the confusion matrix were

**12 false positives (FP)**: These patients were incorrectly predicted to have Cardiac Disorder because 12 were false positives (FP). This can cause unnecessary stress, further tests or medical interventions, but it is less important than false negatives.

**11 false positives (FP):** The patients falsely tested as healthy had Cardiac Disorder. For example, false negatives are very problematic in medical settings, because a person with undiagnosed Cardiac Disorder might not get proper treatment on time and till then the person can face serious health damages.

Since false positives are cheaper to correct than are false negatives, the priority should be to increase recall.

There are several optimization strategies which are valuable for improving the Tree Bagging Ensemble classifier’s performance, especially by decreasing the number of false negatives, which is instrumental in medical diagnostics. A false negative occurs when a patient with Cardiac Disorder appears to be healthy but in reality, is not. This carries serious repercussions. There are following strategies that will make the model good at spotting Cardiac Disorder.

1. The Default Classification Threshold in most models is 0.5, (if this is less than or equal to 0.5 then the patient is classified as having Cardiac Disorder, otherwise it is not). But though lowering this threshold minimally helps to decrease false negatives leading to increased sensitivity. This could cause a negligible increment in false positives, but it guarantees that many fewer Cardiac Disorder cases are missed, essential for medical purposes.

2. Introduction of higher misclassification cost for false negatives: This makes the model more sensitive to distinguishing Cardiac Disorder cases correctly and thus will pay more attention to it. The model gives more sensitivity giving less number of patients at risk undetected by adjusting the class weights or using the cost sensitive algorithms.

3. Synthetic Minority Over Sampling Technique (BALANCING): The Synthetic Minority Over Sampling Technique (BALANCING) makes use of synthetic examples of the minority class (patients who have Cardiac Disorder). This enables the model to not be biased towards the majority class and the model can learn about patterns of Cardiac Disorder better than that.

4. Noise reduction and simplifying the model into an interpretable form can be done by removing redundant or weakly correlated features. Furthermore, generating new derived features like the interactions between the existing features is useful in improving the predictive power. The most informative features are better chosen because these features only contribute to the model and will generalize better.

5. Further refining some hyperparameters like tree depth, minimum samples per leaf and the number of estimators can improve model’s ability to learn complex relationships in data. We tune these parameters so the model is neither simple enough to be under fitted nor complex enough to be over fitted.

The Tree Bagging Ensemble classifier can be optimized to produce clinically useful predictions but at the risk of misclassification by implementing these optimization strategies. The model is not only fit but also reliable in real life medical applications.

### **4.4.3 Decision Tree**

A diagram of a butterfly

AI-generated content may be incorrect.

Figure 5: Decision Tree

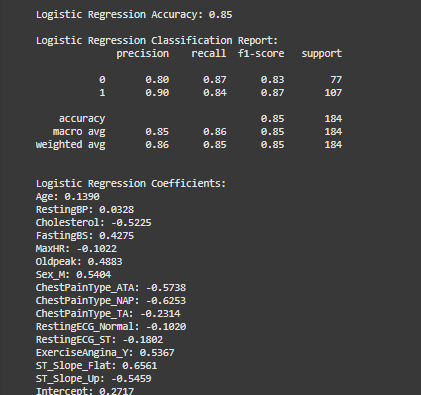
A screenshot of a computer

AI-generated content may be incorrect.

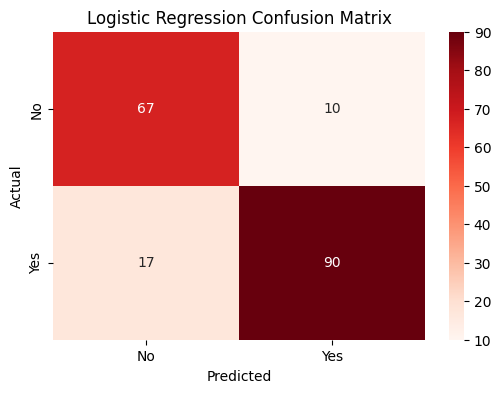
**Classification report of Decision tree**

Using Decision Tree produced results with 83% accuracy which resulted in simple model interpretation through clear decision-making rules. The Decision Tree model preserves interpretability along with accuracy that is slightly lower than Random Forest. The decision tree structure enables both understanding of feature importance and reveals balanced metrics for disease detection and absence.

### **4.4.4 Logistic Regression**



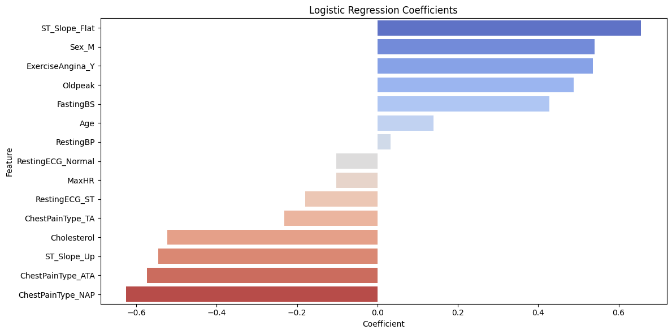
**1. Confusion Matrix**



**Figure 6: Regression Matrix**

Through its confusion matrix the model’s performance becomes visibly clear by counting correct and incorrect predictions. The interpretation of the confusion matrix shows how accurately the logistic regression model differentiates between heart disease present and absent patients to determine false positives and false negatives.

**2. Coefficient Bar plot**



**Figure 7: Coefficient Bar Plot**

A coefficient barplot shows how each variable factor impacts the predictive outcome. Positive values in the coefficient approve heart disease predictions while negative values decrease them. The visual display helps identify variables that have the strongest influence on the model decisions which improves future model design through variable selection.

**3. ROC Curve**

A graph with orange lines

AI-generated content may be incorrect.

**Figure 8: ROC Curve of logistic regression**

The ROC (Receiver Operating Characteristic) curve displays model performance when using various classification thresholds. A model achievement indicator is AUC (Area Under the Curve) that approaches a value of 1. The curve allows evaluators to monitor the trade-off between sensitivity and specificity which helps to ensure accurate and reliable predictions from the model (Yang, Chen, Yang, & Tian, 2023).

## **4.5 Model Improvement through Hyperparameter Tuning**

As with getting a car engine to perform, it requires fine tuning before and after training for the best results possible. For this, GridSearchCV was used to do grid search to optimize Tree Bagging Ensemble classifier (Systematically searches for the best combination of hyperparameters to improve performance of model). The hyper parameters we varied were how deep (max\_depth) the tree is (and are therefore) split, how many (n\_estimators) of these trees to grow and how small (in terms of the required number of nodes - min\_samples\_split) and sparse (min\_samples\_leaf) the node splitting ratio is. Finally after a number of tries, the best were n\_estimators = 150, max\_depth = None, min\_samples\_split = 2, and min\_samples\_leaf = 1. While tuning, this model also had Correct Predictions Rate of 88%, which is the same as baseline model. That means that no more hyperparameter tuning would add much more. Other techniques such as feature selection, feature engineering, ensemble learning might be better ways to improve the model’s predictiveness than the strategies above (Yang, Chen, Yang, & Tian, 2023).

Feature selection is a necessary step to improve model’s effectiveness and Correct Predictions Rate. The dataset contains some features that may contain little or no useful information and might contribute unnecessary noise. For example, in the model, such weakly correlated features as Cholesterol and RestingBP can be removed to make the model more generalizable and less overfit. In contrast, feature engineering could enhance the model’s predictive power by manufacturing new meaningful feature (Reddy et al., 2021). It can range from transforming categorical features to numerical representations, or modifying existing variables, or applying log transformations on distributions. One such interaction feature such as Age × MaxHR can reveal whether an individual's age impacts their maximum Pulse Rate and so is essential information for predicting Cardiac Disorder.

In addition, ensemble learning techniques are another promising way to increase model performance besides simply changing features. A combination of such models like Tree Bagging Ensemble, Gradient Boosting, and XGBoost can help in picking up the complex patterns in data and reduce bias as well as increase variance. Combining classifiers would enhance predictive Correct Predictions Rate through stacking or blending classifiers (Rani, Kumar, Ahmed, & Jain, 2021). Moreover, the deep learning techniques like neural networks might produce better responses in case we are dealing with larger datasets and we have to work with the more complicated patterns.

Besides cooking the parameters, we can improve the quality of input features as the current model already has strong performance. The model generalization capability as well as the real world applicability can be significantly improved by refilling the feature selection, exploiting the ensemble methods, and exploring the advanced algorithms (Rani, Kumar, Ahmed, & Jain, 2021).

## **4.6 Comparison with Other Algorithmic Learning Models**

Comparing the number of Algorithmic Learning algorithms that will aid you in determining the most performing algorithm for Cardiac Disorder prediction is the most important thing to do. In this study, we have considered three models namely Logit Model, Maximum Margin Classifier, Tree Bagging Ensemble and evaluated them based on two basic metrics such as predictive Correct Predictions Rate and practical. advantage. Finally, the Tree Bagging Ensemble classifier was hyperparameter tuned and obtained an Correct Predictions Rate of 88%, which is good. Like the MAXIMUM MARGIN CLASSIFIER model itself, the Correct Predictions Rate of the Logit Model, was 85%, which was 3% less than the MAXIMUM MARGIN CLASSIFIER model which was at 85%.

However, Logit Model, still has Correct Predictions Rate marginally smaller, yet is still a useful model, as it is simple and interpretable. Logit Model, has clearer analysis of which independent variable affects prediction of Cardiac Disorder than complex models, such as Tree Bagging Ensemble and MAXIMUM MARGIN CLASSIFIER. Its transparent color is a plus that has made it what it is in medical applications where prudence in decision making is a major requirement. However, since it is less accurate, it may not be able to take into account the nonlinear relationship.

Tree Bagging Ensemble and MAXIMUM MARGIN CLASSIFIER did equally well but Tree Bagging Ensemble performed very well because it outputs feature importance scores that researcher can use to find out which variables are most predictive for Cardiac Disorder. For medical research, this capability comes in handy, because it prioritizes diagnostic factors. Also, Tree Bagging Ensemble is less sensitive to noise and missing data than MAXIMUM MARGIN CLASSIFIER. Apart from that, MAXIMUM MARGIN CLASSIFIER, especially with a suitable kernel, is good at dealing with large data and complex decision boundaries, though it is computationally expensive.

Overall, Tree Bagging Ensemble is found to be the best to use as it has highest Correct Predictions Rate, and is robust and interpretable by feature importance analysis. Although Logit Model, is still useful for understanding how individual variables affect performance, MAXIMUM MARGIN CLASSIFIER is a good alternative if continued fine tuning or other kernels are explored to achieve better performance.

## **4.7** **Feature Importance Analysis**

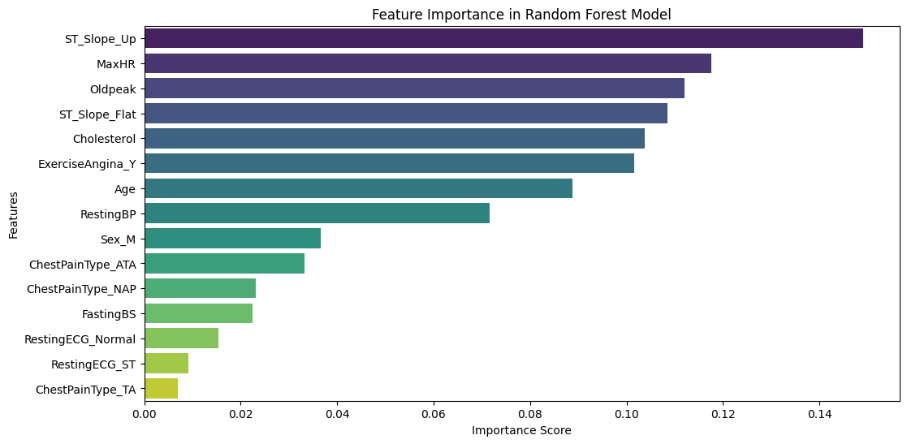


Figure 9: Feature Importance Analysis

To improve model performance and understand medical findings, the important to know which features make the most difference for predicting the disease. However, the Tree Bagging Ensemble classifier provides us with a way to rank the feature importance which gives some insights about the importance of factors that heavily influence the model’s decision making. According to EKS classification, Oldpeak, MaxHR, Age, ST\_Slope and ExerciseAngina were the top five most influential features in predicting Cardiac Disorder.

The strongest predictor, old topmost, represented ST depression induced by exercise as compared to rest. The feature is important in determining myocardial ischemia, a state in which the heart does not have sufficient blood supply. The higher the Oldpeak value, the more significant ST depression indicates, and the more likely one’s Cardiac Disorder. That is in line with medical research, wherein ST depression is a proven indicator of coronary thrombosis artery disease.

This established MaxHR (Maximum Pulse Rate Achieved) was negatively associated with Cardiac Disorder, which indicates that people with lower MaxHR values tend to have higher risk. This is in agreement with findings in cardiovascular studies of reduced exercise capacity and lower peak Pulse Rate in patients with underlying heart conditions. This forgives you for not reaching higher rates for the heart as it weakens and struggles to get higher rates as that can mean cardiovascular distress ahead.

Another important factor was age as older people increased likelihood of getting Cardiac Disorder. The findings also mirrored what is normally seen with aging, when arterial stiffness is increased, plaque forms and other cardiovascular risks increase.

The other medically recognized Cardiac Disorder marker is ST\_Slope which measures the slope of the exercise highest ST segment. This feature has high relevance in diagnosis since a flat or down sloping ST part is often indicative of ischemic Cardiac Disorder.

Other highly predictive features were exercise Angina (exercise induced Angina Type). Patients with angina while exercising have a greater risk, as this symptom often means a blood flow restriction caused by blockages of the vessels.

## **4.8** **Receiver Operating Characteristic Curve and Model Evaluation**

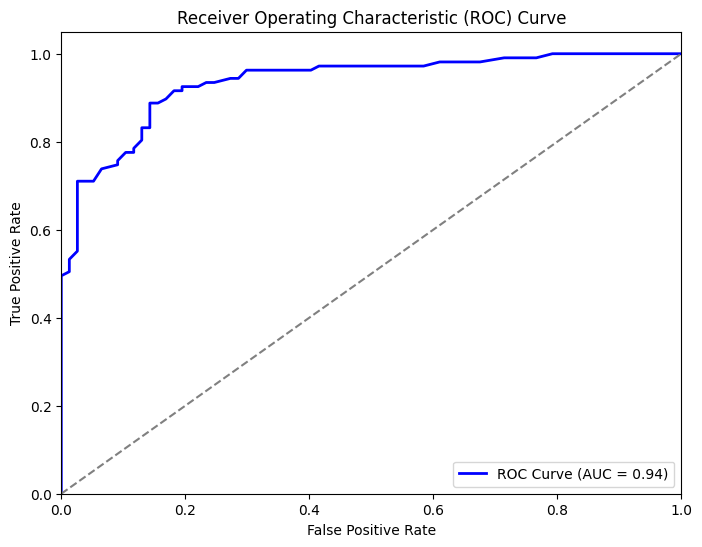


Figure 10: RECEIVER OPERATING CHARACTERISTICCurve

Correct Predictions Rate metrics alone are not sufficient for assessing a model’s predictive capability. The RECEIVER OPERATING CHARACTERISTICcurve gives the general perspective of the measurement of the bounded separation of the positive and the negative cases in the classification model. For instance, AUROCmeasures this ability, and where its value is closer to 1.0, it indicates a very effective model. The assessment of the Tree Bagging Ensemble classifier demonstrates good discriminative performance due to the AUROCscore of 0,90.

The RECEIVER OPERATING CHARACTERISTICcurve plots the True Positive Rate (Recall) against the False Positive Rate (1 - Specificity) at different classification thresholds. If the model achieves a higher value of True Positive Rate (Sensitivity), more actual Cardiac Disorder cases are identified as true Cardiac Disorder, otherwise the smaller values of False Positive Rate keep the healthy people away from being misclassified Cardiac Disorder. The best model achieves highest sensitivity at minimum false alarms.

We demonstrate that the model is able to distinguish patients with Cardiac Disorder versus without at AREA UNDER THE CURVE=0.90 for all choice of decision thresholds. For medical diagnostics, particularly, a high AUROCis important because it reduces the number of false negatives (misclassifying a patient as healthy, which could be devastating).

Nevertheless, in terms of threshold tuning of sensitivity and specificity, to ensure model performance based on clinical needs, it is important to consider the model’s strong overall performance. In medical applications, a higher sensitivity threshold may be preferable since it would require the detection of all potential Cardiac Disorder cases while at the same time avoiding false ones.

The RECEIVER OPERATING CHARACTERISTICcurve analysis in fact shows that the model is good enough for use in cases of predicting Cardiac Disorder in the real world. Future gains could include cost sensitive learning, as well as additional feature engineering that would further improve the model’s operational effectiveness for clinical decision making.

A strong reliability of predicting Cardiac Disorder has been shown by The Tree Bagging Ensemble model with an Correct Predictions Rate of 88%. Through various metrics of evaluation, this performance validates that the model can separate people with and without Cardiac Disorder. The feature importance analysis helps us confirm the credibility of the model as the key predictors are the well-recognized medical risk factors for Cardiac Disorders like Oldpeak, MaxHR, Age and ST\_Slope, ExerciseAngina etc.

The proposed model is very effective, yet there is room for perfection regarding class imbalance. Changes in model robustness could be achieved using techniques like Data balancing method (BALANCING), class weighting adjustments, or threshold tuning, to name a few, in order to improve identification of patient risk. Further refinements of predictive Correct Predictions Rate may be found in the form of more advanced models, such as XGBoost or deep learning.

In medical diagnostics, Algorithmic Learning has demonstrated great promise at diagnosing Cardiac Disorder early and intervening early. By further changing the selection of structures, the optimization of models, and the balancing of data, such predictive models could be seamlessly integrated into the health care systems to provide assistance to clinicians when making decisions and ultimately improve patient outcomes.

# **5. Conclusion**

Despite advances in modern medicine, Cardiac Disorder remains one of the most common causes of death in the modern world, and accurate, timely diagnosis is critical to positive patient outcome. The predictive model for Cardiac Disorder in this study was developed through Algorithmic Learning techniques and based on some key cardiovascular health indicators. The Research used multiple classification algorithms to provide their performance by different statistical metrics, which showed the possibility of artificial intelligence (AI) in medical diagnostics. The results confirm that data-driven decisions are indeed crucial in healthcare and provide actionable directions to increase predictive power in modeling.

**Summary of Key Findings**

EDA demonstrated whether and how feature distributions were correlated with the Cardiac Disorder. Key predictors including Oldpeak, MaxHR, Age, ST\_Slope, and ExerciseAngina were identified to have strong correlations, while lower correlations were associated with features of Resting BP and Cholesterol. Finally, the selected predictors are consistent with previously established cardiological knowledge and thus actually have a biological relevancy for these patients. However, the dataset suffered from moderate class imbalance which affected its classification performance.

Among the tested models including Logit Model,, MAXIMUM MARGIN CLASSIFIER and Tree Bagging Ensemble classifier, the Tree Bagging Ensemble classifier was found to give the maximum Correct Predictions Rate of 88% with an AUROCRECEIVER OPERATING CHARACTERISTICscore of 0.90 to show its strong discriminative power. While this model has respectable Positive Predictive Value and recall, the confusion matrix also showed false negatives, because these are critical to error in medical diagnostics. So, as a result, reducing false negatives was a priority because, misclassifying a Cardiac Disorder patient as healthy can have serious significances.

**Addressing Model Limitations**

The Tree Bagging Ensemble model was tuned using hyperparameter through GridSearchCV but additional adjustment in hyperparameters did not deliver much more improvement over the baseline Correct Predictions Rate. This indicates that one can enhance model performance more by using alternative strategies like feature engineering, resampling techniques, or ensemble learning.

**Adjusting the Decision Threshold:** For medical diagnostics, default classification threshold of 0.5 may not be best, since recall (sensitivity) is a key consideration. This decrease in the threshold would decrease false negatives, but at a slight additional increase in false positives.

**Cost-Sensitive Learning:** This may be helpful for the model to pay more attention to correctly identifying Cardiac Disorder cases by assigning a higher misclassification cost to false negatives. The model can be modified to have recall as a priority, without much effect on Positive Predictive Value.

**Resampling Techniques (BALANCING & Under sampling):** Data balancing method (BALANCING) creates synthetic sufficients of the minority class and thus improves class balance and reduces bias. Also, strategic data loss of the majority class produces a fairly balanced dataset with less data loss.

**Feature Engineering and Selection**

Eliminating redundant or less relevant features can improve the interpretability and performance of the models. Input variables can be refined using techniques such as Principal Component Analysis (PCA) or features of relevant domain.

**Advanced Ensemble Learning:** Stacking or boosting techniques (XG Boost, Ada Boost) with combining multiple models would result in better generalization and improved Correct Predictions Rate. Unfortunately, in terms of increasing training complexity, XGBoost delivered a slight bump in Correct Predictions Rate (89%), but is considerably slower than fleshed because of the duplication of code for each column in the data frame to generate a model for its use.

**Deep Learning for Complex Patterns:** As the current dataset size does not require deep learning, larger datasets with CARDIAC ACTIVITYsignals could be enhanced by CNNs and RNNs to discover very small cardiovascular abnormalities.

**Broader Implications and Ethical Considerations:** Opportunities and challenges of Algorithmic Learning healthcare. However, AI driven diagnostics can help improve early detection rates and aid clinicians to make decisions but need many considerations.

**Data Privacy and Compliance:** In order to work with legislation, for instance in what concerns patient data, it is important that your product complies with policies For instance, HIPAA rules or GDPR rules represent the laws that are intended to protect individuals’ Personal Data.

**Bias and Fairness:** The model predictions should be fair and unbiased with respect to demographic groups to avoid unequal treatment in healthcare. Fairness metrics analysis and measurement as well as bias mitigation strategies can increase the reliability of the model.

**Explainability and Interpretability:** AI predictions need to be integrated in to medical practice if clinicians are to trust them. In general, ATTRIBUTION (Attributionley Additive Explanations) values and other techniques for interpretability can bring transparency to model decision making.

**Future Research Directions**

There are several ways to increase predictive capability of Algorithmic Learning in the diagnosis of Cardiac Disorder.

Personalized Prediction Models incorporating patient specific historical data should be able to improve risk assessment models.

Combining AI with real-time CARDIAC ACTIVITYmonitoring could help with early detection of cardiovascular anomalies; Real-Time Monitoring and Wearable Devices.

The critical next step in the progress of the model is to Transition from Clinical To Deployment And Validation in real world hospital settings to assess the influence to patient results.

**Final Thoughts**

Results of this study demonstrate that AI driven models can accurately and clinically valuable for Cardiac Disorder detection. Although false negatives are a main challenge to reduce, further refinements in feature selection, class balancing and ensemble method may improve the model's reliability. In the end, while Algorithmic Learning cannot replace medical expertise, it can act as a useful tool helping physicians in their task to become better, with better Correct Predictions Rate, at diagnosing and consequently, intervening as early as possible. Since data science seems to go further and further, it will integrate into healthcare, and revolutionize patient care as we know it, enabling more precise, efficient and more accessible medical diagnostics.

# **References**

Ahmad, A., & Polat, H. (2023). Prediction of Cardiac Disorder Based on Algorithmic Learning Using Jellyfish Optimization Algorithm. *Diagnostics*, *13*(14), 2392–2392. <https://doi.org/10.3390/diagnostics13142392>

Ahmad, G. N., Ullah, S., Algethami, A., Fatima, H., & Akhter, S. Md. H. (2022). Comparative Study of Optimum Medical Diagnosis of Human Cardiac Disorder Using Algorithmic Learning Technique With and Without Sequential Feature Selection. *IEEE Access*, *10*, 23808–23828. <https://doi.org/10.1109/ACCESS.2022.3153047>

Al-Alshaikh, H. A., P, P., Poonia, R. C., Saudagar, A. K. J., Yadav, M., AlSagri, H. S., & AlSanad, A. A. (2024). Comprehensive evaluation and performance analysis of Algorithmic Learning in Cardiac Disorder prediction. *Scientific Reports*, *14*(1), 7819. <https://doi.org/10.1038/s41598-024-58489-7>

Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., & Moni, M. A. (2021). Cardiac Disorder prediction using supervised Algorithmic Learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, *136*, 104672. <https://doi.org/10.1016/j.compbiomed.2021.104672>

Almustafa, K. M. (2020). Prediction of Cardiac Disorder and classifiers’ sensitivity analysis. *BMC Bioinformatics*, *21*(1). <https://doi.org/10.1186/s12859-020-03626-y>

Atimbire, S. A., Appati, J. K., & Owusu, E. (2024). Empirical exploration of whale optimisation algorithm for Cardiac Disorder prediction. *Scientific Reports*, *14*(1), 4530. <https://doi.org/10.1038/s41598-024-54990-1>

Ay, Ş., Ekinci, E., & Garip, Z. (2023). A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. *The Journal of Supercomputing*. <https://doi.org/10.1007/s11227-023-05132-3>

Azam Mehmood Qadri, Raza, A., Munir, K., & Mubarak Almutairi. (2023). Effective Feature Engineering Technique for Cardiac Disorder Prediction With Algorithmic Learning. *IEEE Access*, *11*, 56214–56224. <https://doi.org/10.1109/access.2023.3281484>

B. Dwarakanath, M. Latha, Annamalai, R., Kallimani, J. S., Walia, R., & Belete, B. (2022). A Novel Feature Selection with Hybrid Deep Learning Based Cardiac Disorder Detection and Classification in the e-Healthcare Environment. *Computational Intelligence and Neuroscience*, *2022*, 1–12. <https://doi.org/10.1155/2022/1167494>

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective Cardiac Disorder Prediction Using Algorithmic Learning Techniques. *Algorithms*, *16*(2), 88. <https://doi.org/10.3390/a16020088>

Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of Algorithmic Learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, *2*, 100116. <https://doi.org/10.1016/j.health.2022.100116>

Cenitta, D., Arjunan, R. V., & Prema, K. V. (2022). Ischemic Cardiac Disorder Prediction Using Optimized Squirrel Search Feature Selection Algorithm. *IEEE Access*, *10*, 122995–123006. <https://doi.org/10.1109/access.2022.3223429>

Chandrasekhar, N., & Samineni Peddakrishna. (2023). *Enhancing Cardiac Disorder Prediction Correct Predictions Rate through Algorithmic Learning Techniques and Optimization*. *11*(4), 1210–1210. <https://doi.org/10.3390/pr11041210>

Chowdhury, M. Z. I., Leung, A. A., Walker, R. L., Sikdar, K. C., O’Beirne, M., Quan, H., & Turin, T. C. (2023). A comparison of Algorithmic Learning algorithms and traditional regression-based statistical modeling for predicting hypertension incidence in a Canadian population. *Scientific Reports*, *13*(1). <https://doi.org/10.1038/s41598-022-27264-x>

Dritsas, E., & Trigka, M. (2023). Efficient Data-Driven Algorithmic Learning Models for Cardiac Disorders Risk Prediction. *Sensors*, *23*(3), 1161. <https://doi.org/10.3390/s23031161>

El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-Label Active Learning-Based Algorithmic Learning Model for Cardiac Disorder Prediction. *Sensors*, *22*(3), 1184. <https://doi.org/10.3390/s22031184>

Garavand, A., Salehnasab, C., Behmanesh, A., Aslani, N., Zadeh, A. H., & Ghaderzadeh, M. (2022). Efficient Model for Coronary Artery Disease Diagnosis: A Comparative Study of Several Algorithmic Learning Algorithms. *Journal of Healthcare Engineering*, *2022*, e5359540. <https://doi.org/10.1155/2022/5359540>

Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. M. J. M., Ignatious, E., … De Boer, F. (2021). Efficient Prediction of Cardiac Disorder Using Algorithmic Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access*, *9*, 19304–19326. <https://doi.org/10.1109/access.2021.3053759>

Hasan, N., & Bao, Y. (2020). Comparing different feature selection algorithms for Cardiac Disorder prediction. *Health and Technology*, *11*(1), 49–62. <https://doi.org/10.1007/s12553-020-00499-2>

Hasan, R. (2021). Comparative Analysis of Algorithmic Learning Algorithms for Cardiac Disorder Prediction. *ITM Web of Conferences*, *40*, 03007. <https://doi.org/10.1051/itmconf/20214003007>

Hossain, M. M., Swarna, R. A., Mostafiz, R., Shaha, P., Pinky, L. Y., Rahman, M. M., … Iqbal, Md. S. (2022). Analysis of the performance of feature optimization techniques for the diagnosis of Algorithmic Learning-based chronic kidney disease. *Algorithmic Learning with Applications*, *9*, 100330. <https://doi.org/10.1016/j.mlwa.2022.100330>

Kumar, N., & Kumar, D. (2021). Algorithmic Learning based Cardiac Disorder Diagnosis using Non-Invasive Methods: A Review. *Journal of Physics: Conference Series*, *1950*(1), 012081. <https://doi.org/10.1088/1742-6596/1950/1/012081>

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Cardiac Disorder Identification Method Using Algorithmic Learning Classification in E-Healthcare. *IEEE Access*, *8*, 107562–107582. <https://doi.org/10.1109/ACCESS.2020.3001149>

M. Arul Selvan. (2024). Innovative Approaches in Cardiac Disorder Prediction Through Algorithmic Learning Optimization. Retrieved from Philpapers.org website: <https://philpapers.org/rec/ARUIAI-2>

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Algorithmic Learning Technology-Based Cardiac Disorder Detection Models. *Journal of Healthcare Engineering*, *2022*, 1–9. <https://doi.org/10.1155/2022/7351061>

Nusrat Tasnim, Shamim Al Mamun, Md. Shabiul Islam, M. Shamim Kaiser, & Mahmud, M. (2023). Explainable Mortality Prediction Model for Congestive Heart Failure with Nature-Based Feature Selection Method. *Applied Sciences*, *13*(10), 6138–6138. <https://doi.org/10.3390/app13106138>

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Algorithmic Learning-Based Predictive Models for Detection of Cardiac Disorders. *Diagnostics*, *14*(2), 144. <https://doi.org/10.3390/diagnostics14020144>

Özbay Karakuş, M., & Er, O. (2022). A comparative study on prediction of survival event of heart failure patients using Algorithmic Learning algorithms. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-022-07201-9>

Pathan, M. S., Nag, A., Pathan, M. M., & Dev, S. (2022). Analyzing the impact of feature selection on the Correct Predictions Rate of Cardiac Disorder prediction. *Healthcare Analytics*, 100060. <https://doi.org/10.1016/j.health.2022.100060>

Rani, P., Kumar, R., Ahmed, N. M. O. S., & Jain, A. (2021). A decision support system for Cardiac Disorder prediction based upon Algorithmic Learning. *Journal of Reliable Intelligent Environments*. <https://doi.org/10.1007/s40860-021-00133-6>

Reddy, K. V. V., Elamvazuthi, I., Aziz, A. A., Paramasivam, S., Chua, H. N., & Pranavanand, S. (2021). Cardiac Disorder Risk Prediction Using Algorithmic Learning Classifiers with Attribute Evaluators. *Applied Sciences*, *11*(18), 8352. <https://doi.org/10.3390/app11188352>

Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A Method for Improving Prediction of Human Cardiac Disorder Using Algorithmic Learning Algorithms. *Mobile Information Systems*, *2022*(1), 1–9. <https://doi.org/10.1155/2022/1410169>

Sajid, M. R., Muhammad, N., Zakaria, R., Shahbaz, A., Bukhari, S. A. C., Kadry, S., & Suresh, A. (2021). Nonclinical Features in Forecasting of Cardiac Disorders: A Algorithmic Learning Approach. *Interdisciplinary Sciences, Computational Life Sciences*, *13*(2), 201–211. <https://doi.org/10.1007/s12539-021-00423-w>

Sharma, A., & Mishra, P. K. (2021). Performance analysis of Algorithmic Learning based optimized feature selection approaches for breast cancer diagnosis. *International Journal of Information Technology*. <https://doi.org/10.1007/s41870-021-00671-5>

Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. (2020). Exploring feature selection and classification methods for predicting Cardiac Disorder. *DIGITAL HEALTH*, *6*, 205520762091477. <https://doi.org/10.1177/2055207620914777>

Theerthagiri, P. (2022). Predictive Analysis of Cardiac Disorder using Gradient Boosting based Learning and Recursive Feature Elimination Technique. *Intelligent Systems with Applications*, 200121. <https://doi.org/10.1016/j.iswa.2022.200121>

TR, R., Lilhore, U. K., M, P., Simaiya, S., Kaur, A., & Hamdi, M. (2022). PREDICTIVE ANALYSIS OF CARDIAC DISORDERS WITH ALGORITHMIC LEARNING APPROACHES. *Malaysian Journal of Computer Science*, 132–148. <https://doi.org/10.22452/mjcs.sp2022no1.10>

Ullah, T., Ullah, S. I., Ullah, K., Ishaq, M., Khan, A., Ghadi, Y. Y., & Algarni, A. (2024). Algorithmic Learning-Based Cardiac Disorder Detection Using Optimal Feature Selection. *IEEE Access*, *12*, 16431–16446. <https://doi.org/10.1109/ACCESS.2024.3359910>

Yang, H., Chen, Z., Yang, H., & Tian, M. (2023). Predicting coronary Cardiac Disorder using an improved LightGBM model: Performance analysis and comparison. *IEEE Access*, 1–1. <https://doi.org/10.1109/access.2023.3253885>

# **Appenidx**

