Heart Failure Prediction using Machine Learning with SHAP

Jeffin George Johnson
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ
jeffinjohnson@email.arizona.edu

Reya Jijy Abraham
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ
reyajijyabraham@email.arizona.edu

***Abstract*—Cardiovascular Diseases (CVD) are one of the leading causes of death worldwide. Therefore, it is important to have a reliable, affordable and accurate system that can diagnose heart disease on time in order to provide efficient treatment before it can lead to severe complications resulting in heart attack. Using major health conditions of patients, various machine learning approaches are used to predict the presence of heart disease. We seek to assess the performance of six machine learning (ML) models: Logistic Regression, Decision trees, K Nearest Neighbor approach (KNN), Support Vector Machines (SVM), AdaBoost and XGBoost, for model prediction. The models were evaluated based on their accuracy with SVM model having the best accuracy. Finally, an explainable approach based on the SHapley Additive exPlanations (SHAP) method is implemented to generate explanations of the model's decisions.**

I. Introduction

Cardiovascular diseases are one of the largest causes of death worldwide, taking an estimated 18 million lives every year, accounting for almost 30% of all deaths worldwide. The majority of deaths are due to heart attacks and strokes, most of them occurring in individuals under the age of 75. Individuals with cardiovascular disease or even those who are at high risk owing to the presence of one or more factors such as diabetes, high blood pressure, hyperlipidemia, cholesterol, etc. require early detection and treatment. This is where a machine learning model can be of great help. Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies. This dataset was created by combining several datasets available independently but not combined before. In this dataset, 5 different heart disease datasets are combined with over 11 common features. The five datasets used for its curation are from Cleveland (303 observations), Hungary (294 observations), Switzerland (123 observations), Long Beach VA (200 observations) and the Stalog (Heart) Data Set (270 observations) providing a total of 1190 observations of which 272 were duplicates. Thus, the dataset used contains records of 918 patients.

These are the features used to predict the risk of heart disease:
i) **Age**: The age of the patient in years.
ii) **Sex**: The sex of the patient; either M: Male or F: Female
iii) **ChestPainType**: The classification of chest pain:
• ASY - Asymptomatic • TA - Typical Angina • ATA - Atypical Angina • NAP - Non-Anginal Pain
iv) **RestingBP**: The resting blood pressure in mm Hg.
v) **Cholesterol**: The serum cholesterol in mm/dl
vi) **FastingBS**: The fasting blood sugar is categorized as:
• 1: Above 120 mg/dl • 0: Below 120 mg/dl
vii) **RestingECG**: The resting electrocardiogram results labeled as:
• Normal - Normal • ST(Sinus Tachycardia) - having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of greater than 0.05 mV), • LVH - showing probable or definite left ventricular hypertrophy by Estes' criteria
viii) **MaxHR**: The maximum heart rate attained which is a numeric value between 60 and 202.
ix) **ExerciseAngina**: Exercise-induced angina is categorized as: • Y - Yes • N - No
x) **Oldpeak**: ST depression induced by exercise relative to rest.
xi) **ST_Slope**: The slope of the peak exercise ST segment:
• Up - upsloping • Down - downsloping • Flat - flat
xii) **HeartDisease**: The output class stating the presence of heart disease:
• 1 - Heart disease • 0 - Normal

In the context of machine learning, the target variable is the output that is produced after performing the modeling on the given inputs. It could be a binary variable denoting '0' or '1' in the case of classification or a continuous variable in the case of regression. In this study, the target variable is HeartDisease which determines whether an individual is likely to have heart disease based on the input parameters like age, gender, and other test results. In this study, we perform classification using the following models: Logistic Regression, Decision Tree, K Nearest Neighbor (KNN), Support Vector Machine (SVM), AdaBoost and XGBoost. Feature scaling also known as Normalization is a popular method used to normalize the range of independent features of data. In the case of data processing, it is usually performed during the pre-processing stage. In general, for Machine Learning, it is

essential to normalize the features so that no features are arbitrarily large, (centering) and all features are on the same scale (scaling). Algorithms that exploit distances or similarities between data samples, such as K-NN and SVM, are sensitive to feature transformations. So, feature scaling is useful, when solving a system of equations, least squares, etc, where there are significant issues due to rounding errors. The scaled dataset is split into a training and a testing set with 70% of the dataset as the train set and 30% of the dataset for the test set.
Due to the lack of explainability in Machine Learning Models, it is often challenging to explain why certain predictions are made, which creates limitations in the medical field. To solve this drawback, this project pools together the Machine Learning algorithms with a framework based on SHapley Additive exPlanations (SHAP). In addition to improving the accuracy of predicting the risk of heart disease, it provides insightful explanations, thereby helping medical professionals better comprehend the decision-making process for evaluating disease severity and provides opportunities for early intervention. The SHAP value is an important tool in Explainable AI or Trusted AI, which is an emerging development in Artificial Intelligence.

II. RELATED WORK

Research in the area of medical diagnosis have made massive strides in the past two decades. Prediction of heart disease using machine learning techniques has been a work in progress for several years. Many papers have tried to implement various machine learning techniques to improve heart disease prediction to maximize accuracy to provide prevention and the best path toward diagnosing the disease for patients. The paper by C. Boukhatem *et al* [3] implements MLP, SVM, Random Forest and Naive Bayes to measure the accuracy of prediction from the given dataset and with SVM yielding the best performance with 91% accuracy.

Since traditional machine learning does not provide explainability of the models, SHAP technique first proposed by Lundberg and Lee in [2] was used to give a better outlook of selection of the most important features. In C. Zhing *et al* [4], SHAP was implemented in the XGBoost model to enhance interpretability and human performance. The objective of this paper is to predict the likelihood of heart disease using the six Machine learning Models and to identify the model that provides the best accuracy and then using it as the basis to perform SHAP to explain the model's behavior.

III. METHODS

A. Data Analysis and Visualization:

Data analysis was performed on the dataset to help us analyze and visualize the impact of different features on heart disease. The boxplot in Fig. 1 illustrates the relationship between age and heart disease. The average age of patients without heart disease is 50 years old. The patients with heart disease have an average age of 55 for men and older for women.

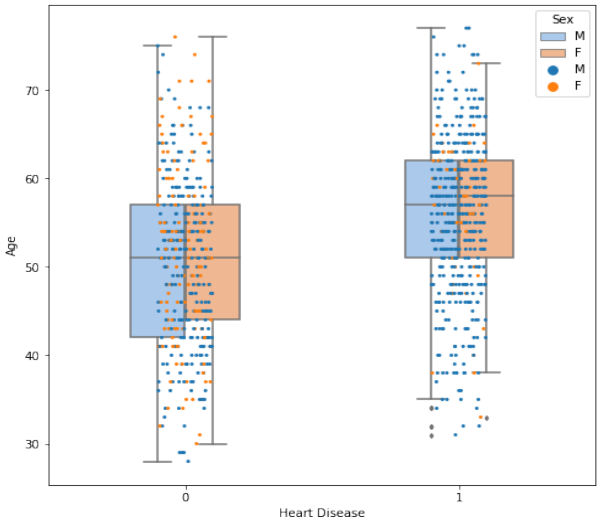


Fig. 1. The relationship between Age and Heart Disease

The violin plot, similar to boxplots, displays numeric data, however, they also display the probability density of the data at different values, usually smoothed by the kernel density estimator. In Fig 2, the relationship between 'Resting BP'(left)/'Maximum Heart Rate'(right) and heart disease is shown. On the right, there is a significant difference in 'MaxHR' between patients who have heart disease and patients who do not. On the other hand, in the chart on the left, there is no significant difference in the level of 'RestingBP' between two types of patients. Fig 3 represents the visualization of

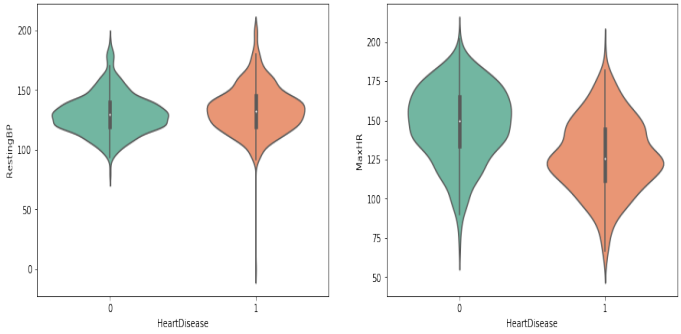


Fig. 2. The relationship between a)Resting BP and heart disease b)Maximum Heart Rate and heart disease

the impact of the features in our dataset. It gives range of the values of Age, Resting BP, Cholesterol, Fasting Blood Sugar, Maximum HR and Oldpeak (the slope of exercise relative to the rest), of the patients who have been diagnosed with heart disease in our Dataset.

B. Data Correlation:

The features in the dataset were correlated to generate a heatmap displayed in Fig 4. The correlation coefficient matrix exhibits the relationship between the various attributes and the output. The correlation close to 1.0 and -1.0 means the

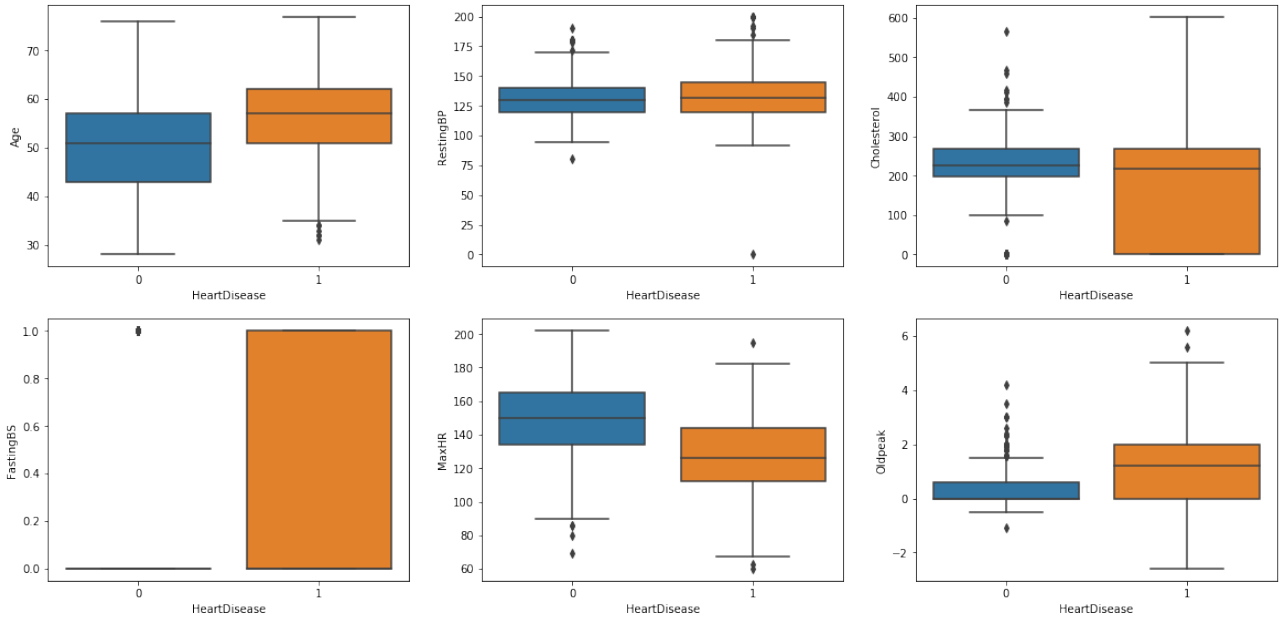


Fig. 3. Feature Visualization

variable is positively and negatively correlated, respectively. The correlation close to 0 is less correlated.

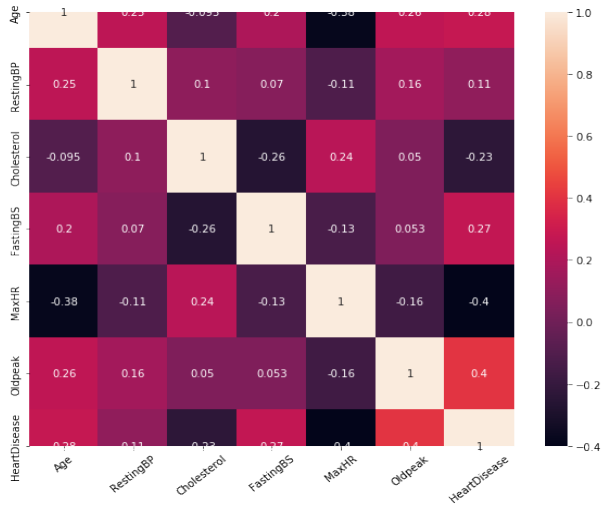


Fig. 4. Heatmap of correlated health factors

C. Data Preprocessing and Cleaning:

Data cleaning is the first step to prepare datasets to ensure that there are no missing values, occurring during the application because most machine learning algorithms cannot work with missing features. Since the dataset is well constructed, there is no need to clean the dataset.

D. Feature Scaling:

Feature Scaling is a preprocessing tool to scale data to a standard range before training machine learning models. Without feature scaling, even an outlier can impact the model. This causes poor performance in the model during learning. The dataset contains both binary and non-binary values. We classify them as continuous and discrete values. The standard deviation of discrete features is small, ranging between 0 and 1, thus there is no need to scale them. However, for continuous features, the standard deviation varies. Hence, scaling these features is important. The features are categorized before performing feature scaling. The dummy variable operation was performed on categorical variables that have more than two categories and was represented by a set of dummy variables with one variable for each category. This variable takes values of 0 and 1, where the values indicate the presence or absence of something. We perform the standardization by using **MinMaxScaler()** to scale features. The scaled data lies between 0 and 1 and the standard deviation of every feature is in the range of 0 and 1 as well. The dataset is now ready to be used in building models.

E. Data Splitting:

In machine learning, the dataset is split into training and testing sets, where the training set is used to train the model, and the testing set is to test it and make predictions on the output. The scaled dataset is split with 70% of the dataset as the training set and 30% of the dataset for the testing set.

F. Algorithms:

1) Decision Tree: Decision tree (DT) is a supervised learning algorithm, mostly used for classification problems. It can deal with both categorical and numerical data. It is a tree-like structure and consists of internal nodes, leaf nodes, and branches. The tree is trained using the training data until all leaf nodes are pure, meaning there is no ability to split

leaf nodes. The maximum depth of the tree is 14.

2) Logistic Regression: Logistic regression is a classification algorithm used for binary classification problems, to predict the value of the predictive variable y when y € [0, 1], 0 represents the negative class and 1 is the positive class. Logistic regression is extensively used to examine data involving a dependent, dichotomous, or binary outcome variable against the independent variable. Normality data with equal variance and covariance for all variables are not needed to execute logistic regression.

3) Support Vector Machine: SVM is a supervised learning technique widely used for classification, regression, and outlier detection. It aims to form a decision boundary between various classes and to label predictions using one or more feature vectors. In order to perform classification, the support vector machine technique finds a hyperplane and maximizes the margin to differentiate between classes. These data points are referred to as support vectors. It deals with both linear and non-linear datasets. SVM can play a major role in prediction of heart diseases. SVM uses a maximum margin strategy that transforms into solving a complex quadratic programming problem. Due to the good performance of SVM in classification, it is used in various applications.

4) K-Nearest Neighbor: K-Nearest Neighbor (KNN) method is one of the simplest machine learning algorithms. It is used for solving both regression and classification problems where there is limited knowledge about the data distribution. It is a supervised learning algorithm. It finds 'k' nearest data points, for which the target value is to be found. Then, it assigns the average of those 'k' data points to the particular data point. The K-NN algorithm predicts the class label of a new input and it utilizes the similarity of new input to its already existing input samples in the training set.

5) AdaBoost: Adaboost, also known as adaptive boosting algorithm, uses the concept of boosting which is an ensemble technique to increase the performance of weak learners. Each copy of the classifier is trained on a different subset of data. Multiple subsets of the dataset are created by assigning weights to data items. If an instance is misclassified, then the weight on that instance is increased, otherwise, the weight is decreased. In this way, multiple models are trained one after another consecutively. After that, these weak classifiers are combined using a cost function to produce a strong classifier. Classifiers with greater accuracy are given higher weightage in the final prediction. The weak classifier to which boosting is applied is passed as a parameter to the algorithm. The default weak classifier used for boosting in Adaboost is a decision tree.

6) XGBoost: XGBoost (Extreme Gradient Boosting) is an ensemble tree method that makes use of the gradient descent

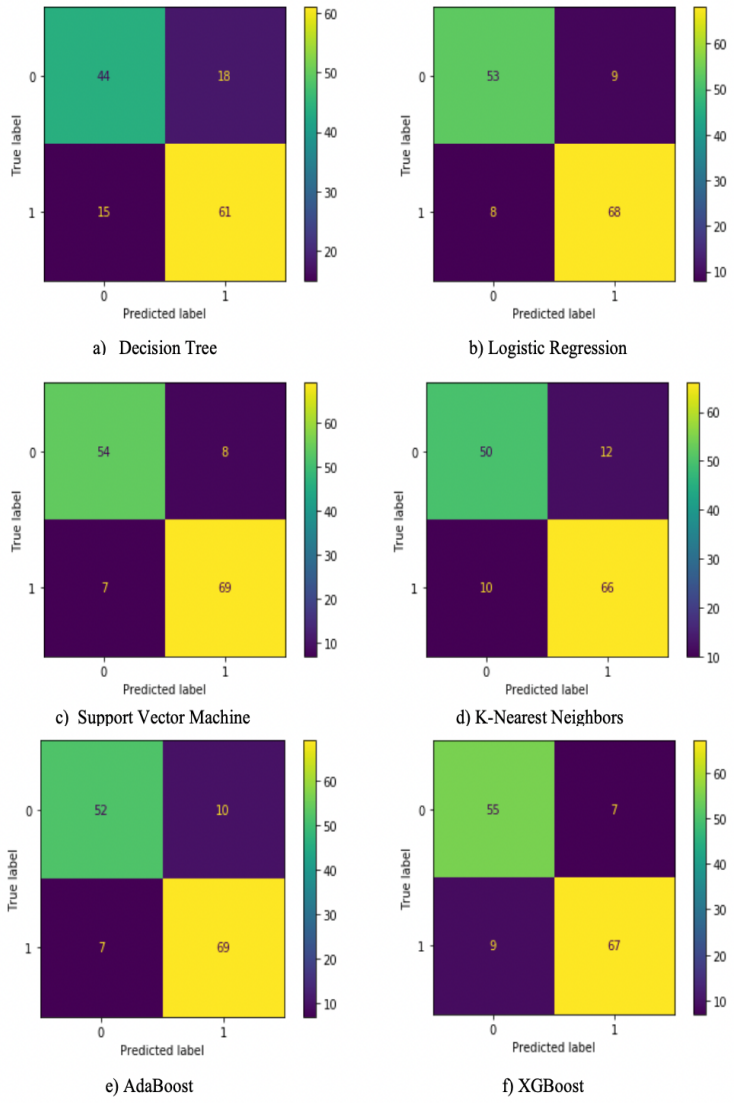


Fig. 5. The Confusion Matrices of the various Models

architecture to boost weak learners. XGboost is a package that belongs to the Distributed Machine Learning Community (DMLC). In XGBoost, first, a weak classifier is fit to the data. It fits one weaker classifier to improve the performance of the current model, without making changes to the previous classifier, and this process continues. Each new classifier has to consider where the previous classifiers were not performing. The predictive power of the XGBoost model is controlled by a loss function and the simplicity of the model is controlled by regularization.

G. Evaluation metrics

Evaluation metrics such Confusion Matrix, Accuracy, Precision, Recall and F1 Score are used to test the quality and performance of any the machine learning model. [3]

1. Confusion Matrix is a square matrix, of size N, where N is the number of classes being predicted. The elements of the matrix are the counts of the correct and incorrect predictions, split by class. A True Positive (TP) is the number of the correctly classified Positive class. In this case, the number of correctly diagnosed heart diseases. Likewise, a True Negative (TN) is the number of correctly classified Negative class. In this case, the count of the accurately predicted absence of heart disease.

2. Accuracy is the percentage of correct predictions out of the total number of data points and is given by the given equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3. Precision is the percentage of the positive cases that were classified correctly, and is obtained from the confusion matrix by the following equation:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4. Sensitivity or Recall is the percentage of the actual positive cases that were classified correctly, and is obtained from the confusion matrix by the following equation:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

5. F1 Score measure is used if the target is to get the best precision and recall, as it provides a harmonic mean of the recall and the precision values in the classification problem, and is obtained from the confusion matrix by the following equation:

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

H. Permutation Importance

Permutation importance calculates the model's prediction error after permuting the features, to measure feature importance. Features are listed in the decreasing order of their importance with the most important feature being listed on the top and the least important feature being listed at the end . In the permutation importance table provided here, the first number shows how much accuracy decreased with a random shuffling and the number after the ± measures how the accuracy has varied from one reshuffling to the next. Negative values are sometimes observed for permutation importance. This happens when the noisy data is more accurate than the real data.

I. Partial Dependence Plot

Partial Dependence plots are used to determine which features contributed the most to a particular classification. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex. The partial dependence of two features can be viewed at once using the PDP interact plot.

TABLE I
RESULTS OF PREDICTED MODELS

	Accuracy	Precision	Recall	F1 Score
Decision Tree	80%	77%	78%	77%
Logistic Regression	88%	87%	85%	87%
SVM	89%	85%	88%	87%
KNN	86%	86%	90%	88%
AdaBoost	84%	85%	85%	85%
XGBoost	84%	85%	86%	85%

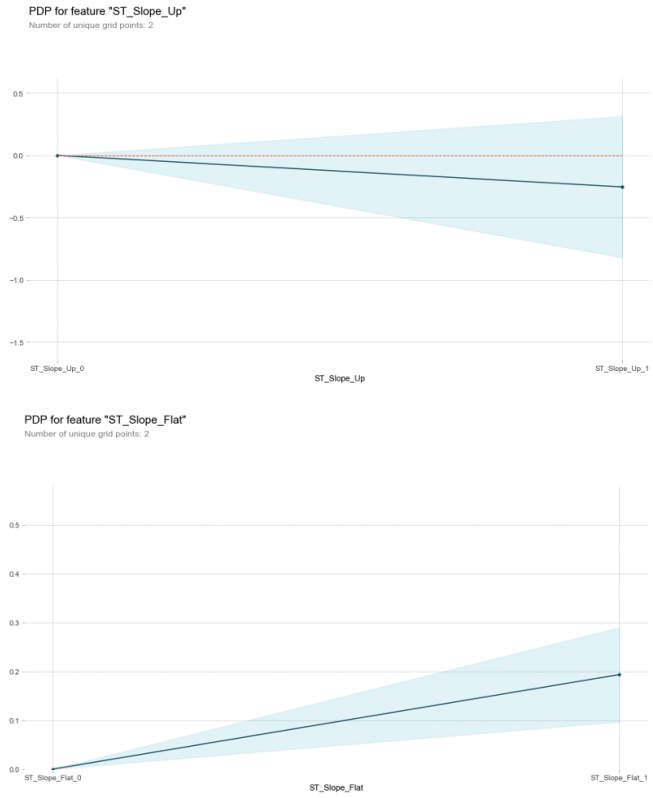


Fig. 6. PDP Plot for ST_Slope_UP and ST_Slope_Flat

J. SHapely Additive exPlanations

The SHAP method, used to interpret Machine Learning predictions, is a work, proposed by Lundberg *et. al* to explain the black-box nature of Machine Learning Models. SHAP can perform both local and global interpretability simultaneously and has superior theoretical foundation compared with other methods. SHAP is used in this project to provide an explanation for the predictive model, which includes related risk factors that lead to death in patients with heart disease. We determined the main factor that contributes to heart disease from the features provided in our dataset by performing SHAP on Decision Tree, SVM and XGBoost Models. The **KernelExplainer**, in SHAP, builds a weighted linear regression by using data, predictions, and functions that predicts the predicted values. It computes the variable importance values based on the Shapley values from game theory, and the coefficients from a local linear regression. The drawback

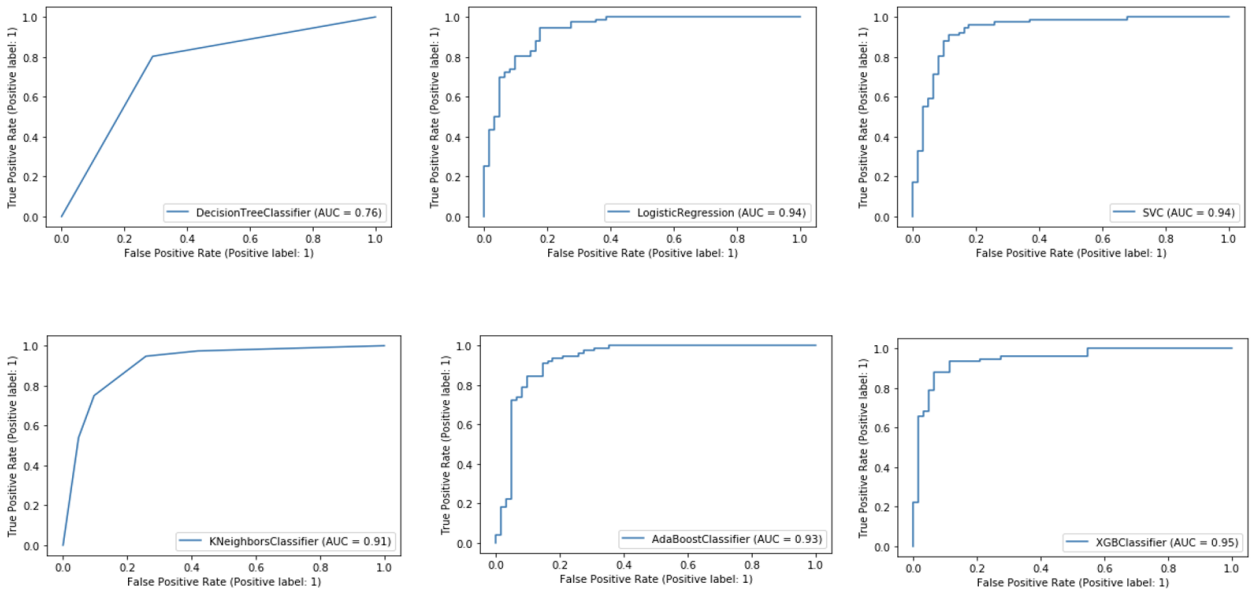


Fig. 7. ROC curves and AUC values of the Machine Learning Models

of the **KernelExplainer** is its long running time. However, in tree-based machine learning models, the **TreeExplainer()**, has been optimized to render fast results. The SHAP Python module does not yet have specifically optimized algorithms for all types of algorithms (such as KNNs). The summary plot, dependence plot and force plot were used to interpret feature importance.

SHAP interaction values are a generalization of SHAP values to higher-order interactions. Fast computation of pairwise interactions are implemented for tree models with SHAP. This returns a matrix for every prediction, where the main effects are on the diagonal and the interaction effects are off-diagonal. These values often reveal interesting hidden relationships.

A ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate and False Positive Rate The precision-recall curve shows the tradeoff between precision and recall for different thresholds. A high area under the curve represents both high recall and high precision, where high precision relates to a low false-positive rate, and high recall relates to a low false-negative rate.

IV. RESULTS AND DISCUSSION

1) Confusion Matrix: The six machine learning techniques were used to build the heart disease prediction model, and the results were obtained to determine the best model. The result of each model based on the confusion matrix is shown in Fig. 5.

2) Evaluation Metrices: According to the evaluation metrices shown in Table I, we can observe that SVM performed best with accuracy of 89%, precision of 85%, recall of 88% and F1 score of 87%.

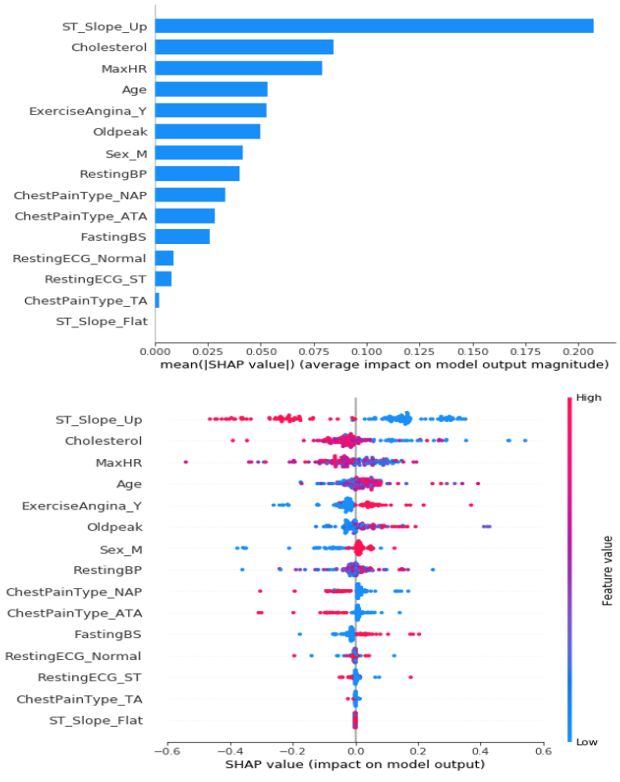
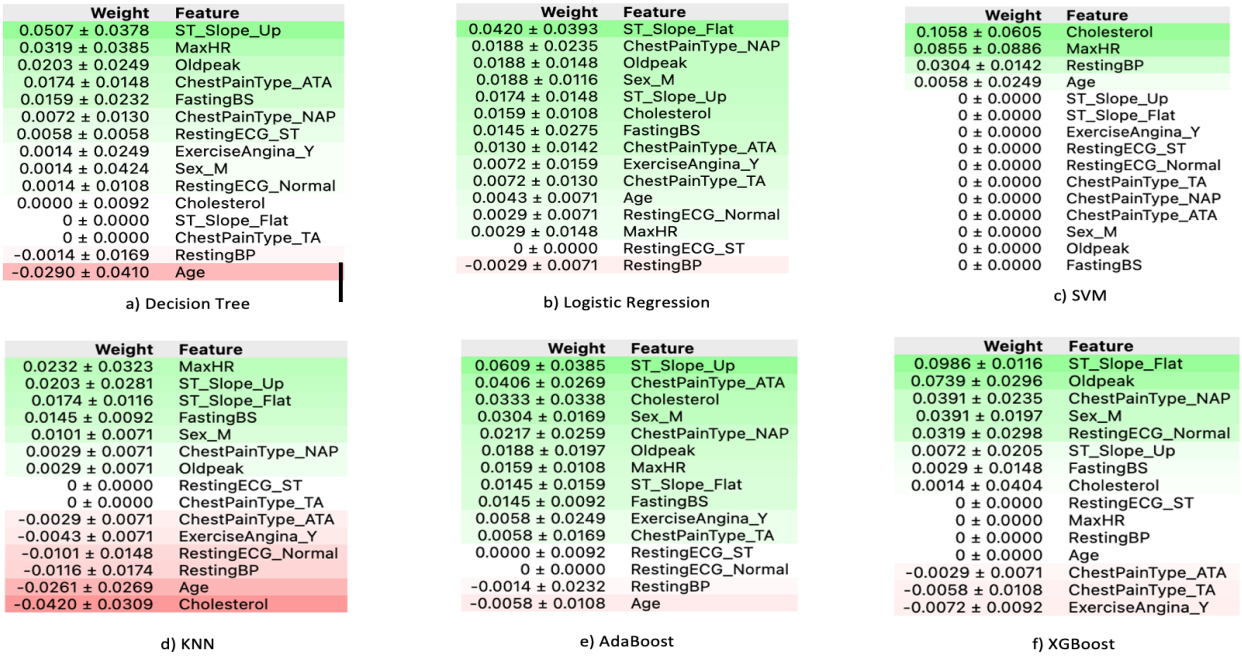


Fig. 8. Summary Plot

3) ROC Curves and AUC: The ROC (Receiver Operating Characteristics) and the AUC (Area under Curve) for the machine learning models is shown in Fig. 7.



Weight	Feature
0.0507 ± 0.0378	ST_Slope_Up
0.0319 ± 0.0385	MaxHR
0.0203 ± 0.0249	Oldpeak
0.0174 ± 0.0148	ChestPainType_ATA
0.0159 ± 0.0232	FastingBS
0.0072 ± 0.0130	ChestPainType_NAP
0.0058 ± 0.0058	RestingECG_ST
0.0014 ± 0.0249	ExerciseAngina_Y
0.0014 ± 0.0424	Sex_M
0.0014 ± 0.0108	RestingECG_Normal
0.0000 ± 0.0092	Cholesterol
0 ± 0.0000	ST_Slope_Flat
0 ± 0.0000	ChestPainType_TA
-0.0014 ± 0.0169	RestingBP
-0.0290 ± 0.0410	Age

a) Decision Tree

Weight	Feature
0.0420 ± 0.0393	ST_Slope_Flat
0.0188 ± 0.0235	ChestPainType_NAP
0.0188 ± 0.0148	Oldpeak
0.0188 ± 0.0116	Sex_M
0.0174 ± 0.0148	ST_Slope_Up
0.0159 ± 0.0108	Cholesterol
0.0145 ± 0.0275	FastingBS
0.0130 ± 0.0142	ChestPainType_ATA
0.0072 ± 0.0159	ExerciseAngina_Y
0.0072 ± 0.0130	ChestPainType_TA
0.0043 ± 0.0071	Age
0.0029 ± 0.0071	RestingECG_Normal
0.0029 ± 0.0148	MaxHR
0 ± 0.0000	RestingECG_ST
-0.0029 ± 0.0071	RestingBP

b) Logistic Regression

Weight	Feature
0.1058 ± 0.0605	Cholesterol
0.0855 ± 0.0886	MaxHR
0.0304 ± 0.0142	RestingBP
0.0058 ± 0.0249	Age
0 ± 0.0000	ST_Slope_Up
0 ± 0.0000	ST_Slope_Flat
0 ± 0.0000	ExerciseAngina_Y
0 ± 0.0000	RestingECG_ST
0 ± 0.0000	RestingECG_Normal
0 ± 0.0000	ChestPainType_TA
0 ± 0.0000	ChestPainType_NAP
0 ± 0.0000	ChestPainType_ATA
0 ± 0.0000	Sex_M
0 ± 0.0000	Oldpeak
0 ± 0.0000	FastingBS

c) SVM

Weight	Feature
0.0232 ± 0.0323	MaxHR
0.0203 ± 0.0281	ST_Slope_Up
0.0174 ± 0.0116	ST_Slope_Flat
0.0145 ± 0.0092	FastingBS
0.0101 ± 0.0071	Sex_M
0.0029 ± 0.0071	ChestPainType_NAP
0.0029 ± 0.0071	Oldpeak
0 ± 0.0000	RestingECG_ST
0 ± 0.0000	ChestPainType_TA
-0.0029 ± 0.0071	ChestPainType_ATA
-0.0043 ± 0.0071	ExerciseAngina_Y
-0.0101 ± 0.0148	RestingECG_Normal
-0.0116 ± 0.0174	RestingBP
-0.0261 ± 0.0269	Age
-0.0420 ± 0.0309	Cholesterol

d) KNN

Weight	Feature
0.0609 ± 0.0385	ST_Slope_Up
0.0406 ± 0.0269	ChestPainType_ATA
0.0333 ± 0.0338	Cholesterol
0.0304 ± 0.0169	Sex_M
0.0217 ± 0.0259	ChestPainType_NAP
0.0188 ± 0.0197	Oldpeak
0.0159 ± 0.0108	MaxHR
0.0145 ± 0.0159	ST_Slope_Flat
0.0145 ± 0.0092	FastingBS
0.0058 ± 0.0249	ExerciseAngina_Y
0.0058 ± 0.0169	ChestPainType_TA
0.0000 ± 0.0092	RestingECG_ST
0 ± 0.0000	RestingECG_Normal
-0.0014 ± 0.0232	RestingBP
-0.0058 ± 0.0108	Age

e) AdaBoost

Weight	Feature
0.0986 ± 0.0116	ST_Slope_Flat
0.0739 ± 0.0296	Oldpeak
0.0391 ± 0.0235	ChestPainType_NAP
0.0391 ± 0.0197	Sex_M
0.0319 ± 0.0298	RestingECG_Normal
0.0072 ± 0.0205	ST_Slope_Up
0.0029 ± 0.0148	FastingBS
0.0014 ± 0.0404	Cholesterol
0 ± 0.0000	RestingECG_ST
0 ± 0.0000	MaxHR
0 ± 0.0000	RestingBP
0 ± 0.0000	Age
-0.0029 ± 0.0071	ChestPainType_ATA
-0.0058 ± 0.0108	ChestPainType_TA
-0.0072 ± 0.0092	ExerciseAngina_Y

f) XGBoost

Fig. 9. Permutation Importance Tables

4) Permutation Importance: From the Permutation importance tables shown in Fig. 9, we can see that ST_Slope is the most important feature as it is an indication of coronary ischemia. The ST segment of a person with regular heart function will have a slight upward concavity. Flat downsloping or depressed ST segments are an indicator of coronary ischemia which eventually leads to heart attacks. Hence it is one of the most important features in predicting the likelihood of heart disease. The high importance of 'Max heart rate' can also be inferred, since this is the immediate, subjective state of the patient at the time of examination as opposed to age, which is a much more general factor, thus having lesser importance.

5) Partial Dependence Plot: PDP plots vary a single variable in a single row across a range of values and see what effect it has on the outcome. It does this for several rows and plots the average effect. Since ST slope was the most important feature as depicted from permutation importance, the PDP plot was observed in Fig. 6. Since the ST slope showing an upward concavity is an indicator of a healthy person, we can observe from the PDP plot that the probability of heart disease decreases as ST_Slope_up increases. Meanwhile, it is observed that the probability of heart disease increases as the ST_Slope follows a flat or down-sloping direction.

In order to visually explain the most important features selected, SHAP is used to illustrate how these features affect the likelihood of heart disease.

6) Summary plot: The Summary Plot gives an interpretation of the model. The bar graph shows the importance of the ranking of the top variables evaluated by the average absolute SHAP value. The optimal model gives the importance of the rankings of the top variables with stability and interpretation. The feature ranking (y-axis) indicates the importance of the predicted model. The SHAP value (x-axis) is a unified index that responds to the influence of a certain feature in the model. The higher the SHAP value of a feature is given, the patient would have a higher risk of heart disease. The red dot represents high-risk value and blue dot represents a low-risk value.

From Fig. 8, we can see that high values of Cholesterol, Age, Old Peak, Exercise Angina and Fasting blood sugar increase the risk of heart disease. Also, it can be observed that the risk of heart disease is higher among men. It also shows that Non-Anginal Pain and Atypical Angina were not major indicators of heart disease.

7) Interpreting Individual predictions: The plots in Fig. 10 provide interpretation of predictions of individual patients.

SHAP values can also show the contribution of each feature for individual patients toward their prediction. In this work, two examples were provided to illustrate the interpretability of the model: (74-year-old male and 63-year-old Female) The arrows show the influence of each factor ion prediction. The blue arrow reduced risk of death and a red arrow indicates an increased risk of death. The final SHAP value corresponds to the prediction score. For a person having a higher SHAP value, the prediction was good and lower SHAP values indicate poor prediction. In this case, the represented man had a higher SHAP value and the woman had a lower SHAP value.

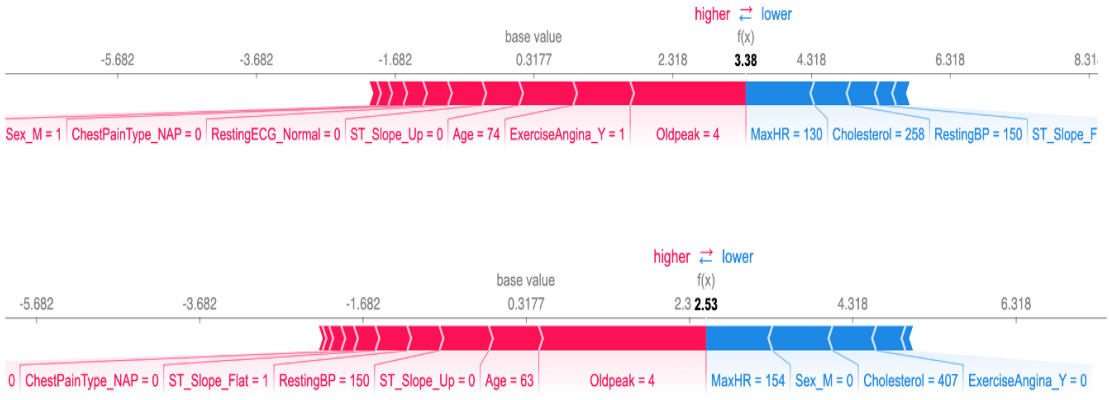


Fig. 10. Individual Force Plot

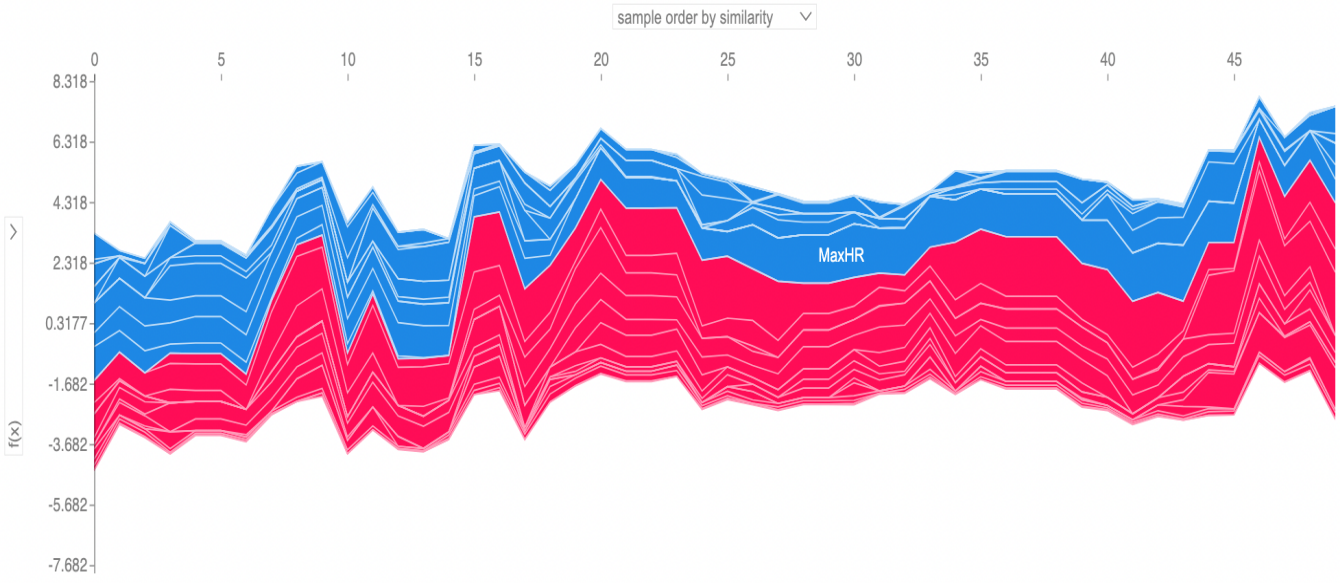


Fig. 11. Collective Force Plot

8) Collective Force Plot: The collective plot shown in Fig. 11 provides a good way to compare and contrast the influence of various features on the likelihood of heart disease for the entire dataset. Since it is interactive, we can hover over the plot to see the reason why each person was predicted to have heart disease(Red) or no heart disease(Blue).

V. CONCLUSION

The use of Machine Learning and SHAP together provides an precise explanation of individualized risk, allowing medical professionals to intuitively understand the influence of key features in the model, and thus helping the decision-making process more efficient for disease severity assessment. This in turn helps to address the root of the problem before the condition of a patient deteriorates severely.

This work demonstrated six learning methods to construct the prediction model. The collected data from Kaggle was cleaned of missing values and outliers. The model was split into training and testing sets and tested for each machine learning algorithm. The SVM algorithm has the best result with 89% accuracy, 85% precision, 88% Recall and 87% F1 score. The SHAP method was then deployed to explain feature importance and generate explanations of the model's decision. It was observed that ST_Slope, Maximum Heart Rate and OldPeak had the highest effect on the likelihoood of heart disease.

This work can be further expanded to include deep learning, to construct medical models. Furthermore, this study included only structured data. This can be extended to include unstructured data along with integrating other relevant clinical risk indicators such as diet, living habits, environmental and other

factors to improve predictions.

REFERENCES

[1] S. M. Lundberg, S. Lee, 'A Unified Approach to Interpreting Model Predictions', 'Advances in Neural Information Processing Systems 30 (NIPS 2017)'.

[2] C. Boukhatem, H. Youssef, and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), vol. Dubai, UAE, Feb 2022.

[3] K. Wang, J. Tian, C. Zhing, H. Yang," Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP, 'Computers in Biology and Medicine Taiyuan,':China, 2021.

[4] A. Nassif, O. Mahdi, Q. Nasir, M. Abu Talib, and M. Azzeh, "Machine Learning Classifications of Coronary Artery Disease." Jan. 2018.

[5] L. Riyaz, M. A. Bhatt, M. Zaman, O. Ayob "heart Disease Prediction using Machine learning Techniques: A quantitative review," 29 August 2021', 'AISC Vol 1394', 'pg 81-94'.

[6] R. Nicole, "Heart Failure Prediction Dataset,'Kaggle.'https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction' (accessed April. 10, 2022).