Machine Learning Empowered by XAI for Cardiac Events Prediction

Kruthi S B
Department of AIML
New Horizon College of Engineeiring
Bengaluru,India
kruthi.s.banakar@hotmail.com

Nithya Shree R
Department of AIML
New Horizon College of Engineeirng
Bengaluru,India
nithyashreeR203@gmail.com

Nitisha Patil
Department of AIML
New Horizon College of Engineeiring
Bengaluru,India
npatil.2021@gmail.com

M Karthik Kumar
Department of AIML
New Horizon College of Engineeiring
Bengaluru, India
mkarthikkumar06@gmail.com

Malem Krishna Vamsi
Department of AIML
New Horizon College of Engineeiring
Bengaluru, India
krishnavamsi2002@gmail.com

S Gunasekar
Department of AIML
New Horizon College of Engineeiring
Bengaluru,India
guna.eie@gmail.com

Abstract*— Cardiovascular disease is the most prevalent cause of death worldwide, and early detection of cardiac events is crucial for effective management and prevention of cardiovascular complications. This study is aimed at integration of machine learning techniques and Explainable Artificial Intelligence (XAI) for predicting the likelihood of cardiac events based on the heart disease dataset from the UCI Machine Learning Repository. We conducted rigorous data preprocessing encompassing outlier identification, elimination of inconsistent and duplicate values. Next, we analysed the features and relationship between them using exploratory data analysis (EDA). We then applied various machine learning algorithms, including K-Nearest Neighbours, Logistic Regression, Random Forest, SVM, Decision Trees, AdaBoost, and Gradient Boosting. To find the optimal model, we explored techniques such as hyperparameter tuning and model evaluation. Through thorough evaluation, we found that the Random Forest algorithm performed the best, achieving a classification accuracy of 95%. The interpretation of the model was enriched by the incorporation of XAI techniques. Our findings emphasize the potential of machine learning techniques further amplified by the integration of XAI as a valuable tool for predicting cardiac events, with the capacity to significantly contribute to the early detection and prevention of cardiovascular complications.

Keywords*— Explainable AI, XAI, UCI Machine Learning Repository, Machine Learning, Cardiovascular Diseases, Random Forest

I. Introduction

The term ‘Cardiovascular Disease’ (CVD) is used to refer to potentially fatal disorders that can impair the heart, its blood vessels, and artery’s function. CVDs have been the major cause of mortality worldwide, accounting for 17.9 million deaths every year and nearly a third of these deaths occur in those under 70 years of age. Coronary heart disease is the most common type leading to increased rates of morbidity and mortality worldwide. Heart disease can be caused by a variety of causes, such as inheritance, high blood pressure, high cholesterol, smoking, diabetes, obesity, lack of physical activity, poor diet, and stress. A heart attack normally needs to be treated within an hour of the first sign of symptoms; failure to do so can result in sudden death. Unfortunately, 11,000 heart attacks are incorrectly diagnosed each year. This is a raising cause of concern and needs an immediate and robust solution. Therefore, identification of the disease in people at the highest risk of developing the disease and ensuring that they receive appropriate treatment after diagnosis can prevent premature deaths. Machine learning can be widely used in the medical sector for the diagnosis of various diseases worldwide. It has shown a remarkable performance in predicting it in its early stages. By analyzing the intricate patterns in the data, machine learning can make accurate predictions. While machine learning models demonstrate remarkable predictive capabilities, their inner workings often remain inscrutable, hindering the establishment of trust and understanding among medical practitioners. This challenge has prompted the adoption of Explainable Artificial Intelligence (XAI) techniques, which aim to unravel the 'black box' of machine learning models and provide interpretable insights into their decision-making processes. The dataset was taken from the University of California, Irvine (UCI) Machine Learning Repository which includes several features that can be significantly important for the purpose of making predictions and producing accurate results. The machine learning model is trained to classify the output as discrete classification type, predicting 0 (absence of cardiac disease) and 1 (presence of cardiac disease). The patients can be diagnosed at an early stage for the disease using the model and approach for future medication. The research is aimed towards building a robust model that can combat overfitting, underfitting and can generalize well on the dataset. The novelty in our paper is rooted in the integration of Explainable Artificial Intelligence (XAI) techniques to enhance the interpretability and utility of machine learning models for CVD prediction. By harnessing the power of XAI, we aim to not only construct an accurate prediction model but also to empower healthcare professionals with a profound comprehension of the intricate nuances underlying the disease.

II. Related Work

This paper has gone through many other research based on heart attack prediction and analysis. They are addressed here.

In the year 2021, Ali et al [1] reported that they initially preprocessed data using WEKA software. Exploratory Data Analysis (EDA) was carried out. The algorithms applied experimented were LR, ABM1, MLP, DT, KNN and Random Forest. These algorithms were compared based on their effectiveness and confusion matrix and feature importance scores were obtained. It is found by 10-fold cross validation that KNN, RF and DT were the best performing algorithms. Also, they found that chest pain (cp) was the most important feature shown by LR, DT, and RF. The process by which the models reached their conclusions was not adequately elucidated.

In the year 2022, as reported by Prusty et al [2], nine different machine learning and deep learning models were used to predict coronary heart disease. The logistic regression classifier gave the best results, with an accuracy of 90.78%. The authors face a challenge in accurately measuring the performance of prediction model as there is no gold standard for diagnosing the disease.

In the year 2021, M. Kavitha et al [3] reported to have used a hybrid model of Decision trees and Random Forests for the prediction. They found that the hybrid model achieved highest accuracy of 88%. The assessment of precision and recall curves, as well as feature scores, was not carried out.

In the year 2021, Nissa et al [4] reported the heart disease prediction using Machine learning algorithms. The authors used comparative algorithms such as SVM, Decision, Random Forest, and hybrid techniques. The research concluded that Random Forest is the best technique for predicting disease. The feature importance and the analysis of the model's predictions were not put forth.

In the same year, Nawaz et al [5] reported the cardiovascular disease prediction empowered with Gradient Descent Optimization (GDO). Proposed models consist of the first phase where the data is collected, pre-processed, and GDO is performed. This result was sent to the validation phase where the model is retrained with GDO. This model is found challenging to interpret.

In the year 2022, Manjula P et al [6] presented the prediction of heart attacks in the year 2022 utilizing machine learning methodologies, algorithms, and techniques such Decision Trees, Logistic Regression, SVM, Naive Bayes, Random Forest, KNN, and XG Boost Classifier. The study's conclusion highlighted the achievable accuracy of 90.16% using the Random Forest algorithm. Nonetheless, the investigation did not encompass the proposition of interpretability for their model.

S. Manikandan [7] used the Gaussian Naïve Bayes algorithm which achieved an accuracy of 81.25%. The author created a web interface that allows users to easily access the classifier. The prototype of the system includes a response variable which can be used to classify individuals on their risk factor for narrowing blood vessels. The authors faced challenges in obtaining a large and comprehensive dataset of individuals with and without narrowing of blood vessels.

A prominent concern observed across various related studies is the absence of Explainable Artificial Intelligence (XAI) techniques. We also observed the lack of a large and high-quality dataset for prediction. The authors used a dataset of 303 patients, which is relatively small for machine learning and deep learning models. While achieving commendable predictive accuracies, these works often lack the integration of XAI, leading to opacity in model decision-making. Hence, we introduce XAI as a novel approach in our research to bridge this gap, enabling transparent and interpretable machine learning models for enhanced cardiovascular disease prediction.

III. Operational System

An operational system refers to a system that is used to operate on the data, which is the fuel for any Machine Learning model, and includes operations like data collection, data processing, analysis, and prediction.

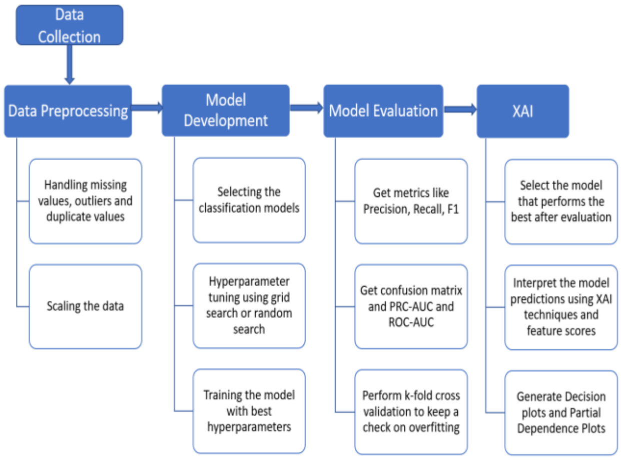

Data Collection
Data Preprocessing
Model Development
Model Evaluation
XAI
Handling missing values, outliers and duplicate values
Scaling the data
Selecting the classification models
Hyperparameter tuning using grid search or random search
Training the model with best hyperparameters
Get metrics like Precision, Recall, F1
Get confusion matrix and PRC-AUC and ROC-AUC
Perform k-fold cross validation to keep a check on overfitting
Select the model that performs the best after evaluation
Interpret the model predictions using XAI techniques and feature scores
Generate Decision plots and Partial Dependence Plots

Fig. 1: Operational System of Machine Learning Technique

A. *Data Collection*

The functioning of the system begins by collecting data from a UCI machine learning repository. The dataset contains 303 rows and 14 columns.

B. *Data Preprocessing*

This involves data cleaning and processing. The dataset has been inspected for duplicate values, null values, and unusual values, also called outliers. The potential outliers were detected using boxplots and were successfully removed by the Interquartile range (IQR) method. With the aim of improving the efficiency of prediction, the dataset also underwent standardization wherever necessary. The dataset was reduced to 299 rows and 14 columns after this process.

C. *Exploratory Data Analysis (EDA)*

The need for analyzing and understanding the data has been satisfied by using various data visualization techniques, which involves production of heatmaps, cluster plots, count plots, boxplots, violin plots, kernel density plots and scatter plots. With the help of these plots, we were able to visualize the correlation between different features and their impact on the output. We could also identify the crucial features that contributed to the prediction of output more accurately.

D. *Data Partitioning*

The dataset was split into subsets for the purpose of training and testing and evaluating machine learning models. The dataset we are dealing with is imbalanced, as one of the classes of the binary output has a greater number of rows than the other. Thus, the use of data splitting techniques like Stratified sampling and Train-test split and Stratified sampling were encouraged. The dataset was divided in a ratio of 80:20 of which 80% was allocated for model training and 20% was reserved for model testing.

E. Preparing the Model

Since we are dealing with a classification problem, we have used supervised machine learning algorithms. This involves the use of algorithms like K–Nearest Neighbors (KNN), Logistic Regression, Decision Trees, Random Forests, Gradient boosting, AdaBoost and SVM. To ensure the model's ability to capture the intricate patterns in the data, we first trained it on the training set. Subsequently, we evaluated its performance on the test set to determine its accuracy. To further enhance the model's performance, we utilized techniques such as Grid Search CV and Random Search to fine-tune its hyperparameters. These methods enable us to find the optimal combination of hyperparameters, leading to improvement of model performance.

F. Final Model Evaluation

The finely tuned model is again tested on the test set and evaluated by the basic performance metrics which includes Accuracy, Precision, Recall and F-measure. In addition, we also analyzed the results using a confusion matrix to better understand the true positives, false positives, true negatives, and false negatives of the model predictions. To ensure that our model generalizes well to new data, we used 10-fold cross-validation and learning curves to assess the performance of the model with different sizes of training data. We also used precision-recall curves (PRC) and receiver operating characteristic (ROC) curves to evaluate the trade-off between precision and recall, and the false positive rate and true positive rate, respectively. The area under the curve (AUC) was calculated to summarize the overall performance of the model across all thresholds.

G. Explainable Artificial Intelligence (XAI)

Once the model is trained and evaluated on the test set, it can be used to make predictions on new, unseen data. We input the features of the new data into the model and obtain the predicted outcome. To enhance the comprehensibility and interpretability of the model's predictions, we undertook the application of SHAP (SHapley Additive exPlanations) plots including summary and decision plots. Furthermore, we employed Partial Dependence Plots (PDPs) to explore the relationships between individual features and predicted outcomes.

IV. Algorithms and Evaluation

A. Algorithms

We use various Supervised Classification Algorithms to build the Machine Learning models. The classifiers are then compared based on their performance and the final model is chosen.

1) K-Nearest Neighbors (KNN):

KNN is a simple supervised machine learning algorithm and can be used for both classification as well as regression. The algorithm works by using ‘k’ nearest data points of the dataset to a new point or the query point and makes predictions based on the majority classes. The model was optimized using Computational geometry methods like Ball Trees. Ball trees are space partitioning data structures as they form binary trees. They form clusters or hyper-spheres (balls) around the datapoints. Each node in the tree represents a ball.

2) Logistic Regression (LR):

LR is a binary classification algorithm that uses the sigmoid function to map any input to a value between 0 and 1. LR uses cross-entropy loss as its cost function, which measures the difference between the predicted probability of the positive class and that of the negative class. The algorithm then uses gradient descent to update the weight vector and bias until convergence, minimizing the cost function. The final predicted output is obtained by applying the sigmoid function to the output of the hypothesis function.

3) Decision Tree (DT):

DT is a well-known machine learning algorithm which uses binary tree to classify data. The Classification and Regression Trees (CART) algorithm is a widely used technique for constructing decision trees. It uses the Gini impurity measure to select the best feature for splitting the data. CART algorithm constructs a binary tree, with each node having two branches: one for samples that satisfy the condition and another for samples that don't.

4) Support Vector Machines (SVM):

SVM is a type of classifier which finds a boundary (or hyperplane) that separates the data points into their respective classes with the largest possible margin. We use the decision function that maps input features to class labels. It returns a score that is used to predict the class of the input. If the score is greater than or equal to 1, the datapoint is classified as positive, and if the score is less than or equal to -1, the datapoint is classified as negative.

5) Random Forest (RF):

A Parallel Ensemble learning technique such as Random Forest builds many models by selecting a random portion of data from the training set, combining those models, and producing results that are superior to those of any one of the individual models alone. The base estimators are strong learners. The algorithm constructs a collection of decision trees followed by aggregation of the predictions of each tree during the training phase to determine the final prediction which is also called bagging.

6) Gradient Boosting (GB):

Gradient boosting is a type of Sequential Ensemble learning model. Sequential ensemble techniques adopt an incremental strategy of merging feeble learners to build a powerful learner by emphasizing errors made in previous iteration. It combines Gradient descent and boosting. Instead of computing the overall true gradient explicitly, gradient boosting aims to approximate the true gradient with a weak learner. This is also called boosting. With a small learning rate, models' prediction is updated slowly, and gradually as new trees are added.

7) Adaboost (AB):

AdaBoost (Adaptive Boosting) is a sequential ensemble model. Initially the model allocates the same weights for all the samples ensuring equal importance to all the samples. In each iteration the decision trees, typically the forest stumps, are tested on different subsets of training data. The model selects the subsets with higher subsets to be used for learning in the next iteration. The first forest stump is created for each feature. The split giving the least misclassified classes is chosen for further splits.

B. *Performance Evaluation Metrics*

1) Basic Metrics:

These are the most used metrics for evaluating classification models, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model's predictions, while Precision and Recall measure the model's ability to correctly identify positive cases and negative cases, respectively. The F1-score is a harmonic mean of precision and recall and provides an overall measure of the model's performance.

2) Cross Validation:

This method is used to evaluate how well a machine learning model generalizes. The dataset is divided into k subsets, and the model is trained and assessed on k-1 subsets while being tested on the final subset. Each subset acts as the test set once during this operation, which is repeated k times. For the classifiers, we carried a 10-fold cross validation.

3) Confusion Matrix:

Confusion Matrix is a table that indicates a classification model's performance. It displays how many True Positives, False Positives, True Negatives, and False Negatives the model predicted. A false negative result could mean that a patient with a serious illness is not identified and treated in a timely manner. In this case, a high recall score (i.e., low False Negative rate) would be desired to ensure that all positive cases are correctly identified, even if this results in a lower precision score.

4) ROC and PRC:

Receiver Operating Characteristic (ROC) curves and Precision-Recall curves (PRC) are two graphical tools used for evaluating the performance of binary classifiers. PRC plots have precision (y-axis) against recall (x-axis), while ROC charts the contrast between True Positive Rate (y-axis) and False Positive Rate (x-axis). The area under the PRC and ROC curves (AUC) are commonly used summary statistics for comparing the performance of different classifiers.

V. Results and Discussion

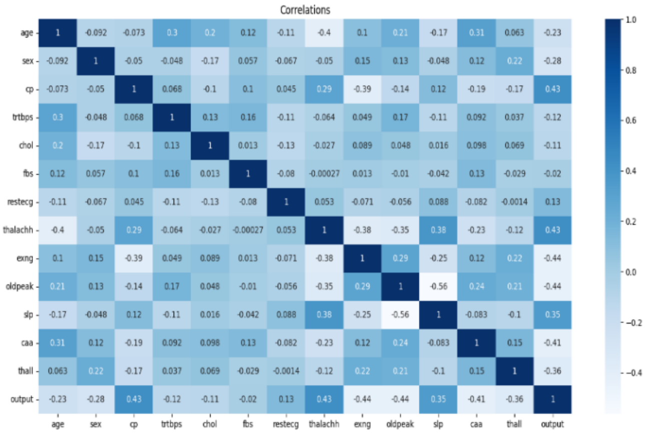
A. *Results of EDA*

The results of Exploratory Data Analysis are being discussed in this section. This helps us to identify important insights, correlations, and inferences from the data, which can later be incorporated into the models.

Fig. 2 shows the correlation between all the features of the dataset coded in heat map. The darker boxes represent higher positive correlation and vice versa. We can infer that restecg, slp, cp, thalachh have positive correlations with output. And, age, sex, thall, caa, exng, oldpeak have negative correlations with output.

Fig. 3 shows Kernel Density Estimation plots for various features. It can be found that people of age group 40 to 60 years have higher probability of having a heart disease contrary to the intuition that elderly people to have higher chances of the disease. The people having maximum heart rate between 150 to 180 beats per min have higher chances of being diseased. Also, the people with chest pain type 0 (typical angina) have lower risk of being diseased.

Fig. 4 represents Violin plots which combine Kernel density plots and box plots. It can be inferred that both women and men in the age group of 50 to 60 years have a higher chance of the disease. Men having cholesterol levels between 200 to 300 mg/dl are at a higher risk of developing the disease. Also, people with heart rates in the range of 140 to180 beats per min are at a higher risk of developing the disease.



Correlations

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
age	1	-0.092	-0.073	0.3	0.2	0.12	-0.11	-0.4	0.1	0.21	-0.17	0.31	0.063	-0.23
sex	-0.092	1	-0.05	-0.048	-0.17	0.057	-0.067	-0.05	0.15	0.13	-0.048	0.12	0.22	-0.28
cp	-0.073	-0.05	1	0.068	-0.1	0.1	0.045	0.29	-0.39	-0.14	0.12	-0.19	-0.17	0.43
trtbps	0.3	-0.048	0.068	1	0.13	0.16	-0.11	-0.064	0.049	0.17	-0.11	0.092	0.037	-0.12
chol	0.2	-0.17	-0.1	0.13	1	0.013	-0.13	-0.027	0.089	0.048	0.016	0.098	0.069	-0.11
fbs	0.12	0.057	0.1	0.16	0.013	1	-0.08	-0.00027	0.013	-0.01	-0.042	0.13	-0.029	-0.02
restecg	-0.11	-0.067	0.045	-0.11	-0.13	-0.08	1	0.053	-0.071	-0.056	0.088	-0.082	-0.0014	0.13
thalachh	-0.4	-0.05	0.29	-0.064	-0.027	-0.00027	0.053	1	-0.38	-0.35	0.38	-0.23	-0.12	0.43
exng	0.1	0.15	-0.39	0.049	0.089	0.013	-0.071	-0.38	1	0.29	-0.25	0.12	0.22	-0.44
oldpeak	0.21	0.13	-0.14	0.17	0.048	-0.01	-0.056	-0.35	0.29	1	-0.56	0.24	0.21	-0.44
slp	-0.17	-0.048	0.12	-0.11	0.016	-0.042	0.088	0.38	-0.25	-0.56	1	-0.083	-0.1	0.35
caa	0.31	0.12	-0.19	0.092	0.098	0.13	-0.082	-0.23	0.12	0.24	-0.083	1	0.15	-0.41
thall	0.063	0.22	-0.17	0.037	0.069	-0.029	-0.0014	-0.12	0.22	0.21	-0.1	0.15	1	-0.36
output	-0.23	-0.28	0.43	-0.12	-0.11	-0.02	0.13	0.43	-0.44	-0.44	0.35	-0.41	-0.36	1

Fig. 2: Heat map of Correlations among all the features of dataset

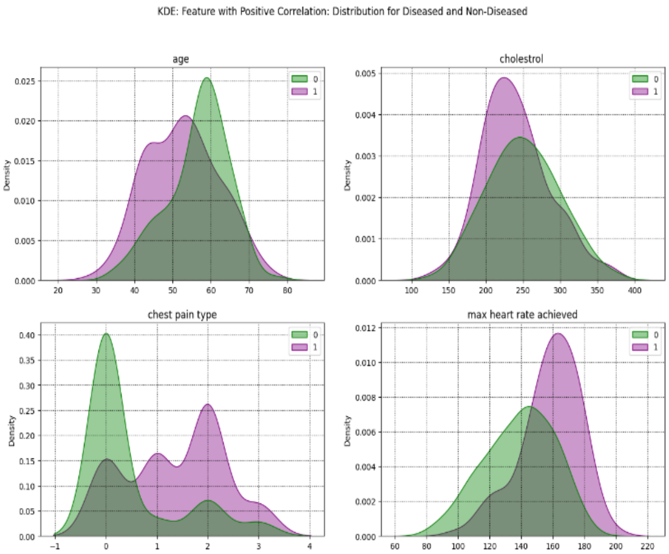


Fig. 3: Kernel Density Estimation plots of selected features.

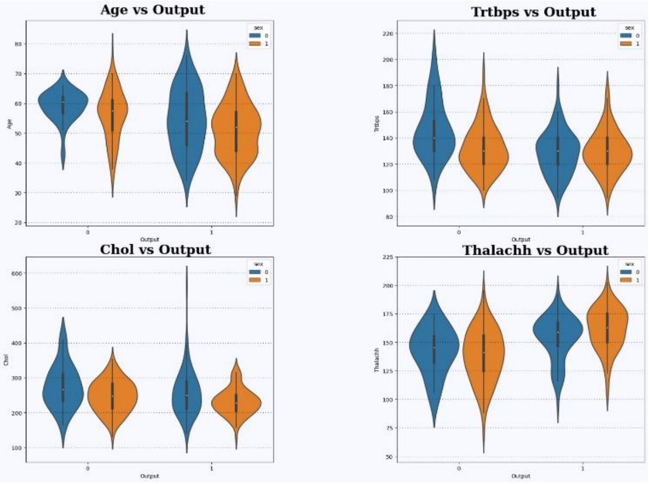


Fig. 4: Violin plots of selected feature.

B. *Results of ML analysis*

This section provides insights on the performance of various classifiers and different aspects for Machine Learning analysis.

Table I illustrates the basic performance outcome parameters of classifications algorithms namely precision, recall, F1 score, accuracy. RF provides maximal accuracy of 95%, a recall of 97%, a precision of 95% and F1 score of 96%.

TABLE I. BASIC PERFORMANCE METRICS

Classifier	Accuracy	Recall	Precision	F1
RF	**0.95**	**0.97**	**0.95**	**0.96**
AB	0.93	0.94	0.94	0.94
GB	0.92	0.94	0.92	0.93
KNN	0.92	0.97	0.9	0.93
SVM	0.92	0.97	0.9	0.93
DT	0.9	0.94	0.89	0.92
LR	0.9	0.97	0.88	0.92

Fig. 5 presents a comparison of the Receiver Operating Characteristic (ROC) results for all seven algorithms under study. Upon examining the ROC curves, it is evident that the Area Under the Curve (AUC) is highest for Ada Boost (0.98), Random Forest (0.97), and Gradient Boosting (0.97). This indicates that these algorithms have a better ability to distinguish between the positive and negative classes compared to the other classifiers.

Fig. 6 presents a comparison of the PR curves and the associated Area Under the Curve (AUC) values for the seven classifiers. Upon examining the PR curves, we find that Ada Boost has the highest AUC value of 0.99, followed closely by Gradient Boosting at 0.98, and Random Forest at 0.97. These values indicate that these algorithms have a better trade-off between precision and recall and are more effective at identifying the positive class than the other algorithms.

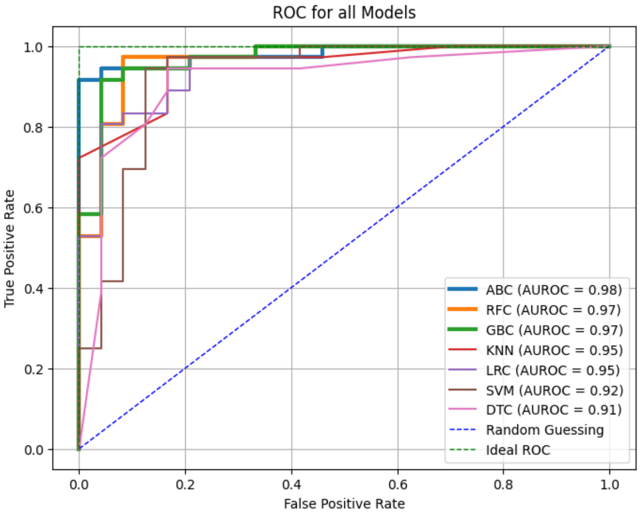


Fig. 5: ROC for all Models

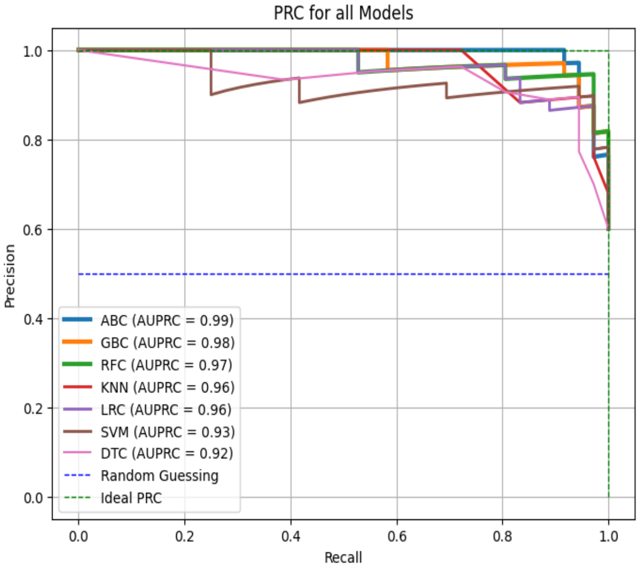


Fig. 6: PRC for all Models

C. *Results of Explainable AI*

This section presents the outcomes and insights derived from the application of Explainable Artificial Intelligence (XAI) techniques, including SHAP (SHapley Additive exPlanations) plots, Partial Dependence Plots (PDPs), and Decision Plots to interpret the predictions from the Random Forest Classifier.

Fig. 7 is the bar plot of summarizing the SHAP values attributed to all features in the Random Forest model. The identification of the top six features, namely cp (chest pain type), caa (number of major vessels colored by fluoroscopy), thall (thalassemia), exng (exercise induced angina), oldpeak (ST depression induced by exercise relative to rest), and thalachh (maximum heart rate achieved), aligns closely with domain knowledge in medical prediction of CVDs.

Fig. 8 presents a decision plot that traces the comprehensive pathway from baseline prediction to the final prediction in our model. This visual representation offers an intuitive glimpse into how individual features collectively influence the model's decision-making process. By tracing this interpretative path, medical practitioners and researchers gain deeper insights into the intricate interplay of features, empowering them with a holistic understanding of the model's decisions.

Fig. 9 displays Partial Dependency Plots (PDPs) illustrating the behavior of the six most influential features in our model. These plots explain how each individual feature influences the predicted outcome while keeping other factors constant. We can infer that categorical feature (cp, caa, thall, exng) exhibit higher relationship with the predicted probabilities.

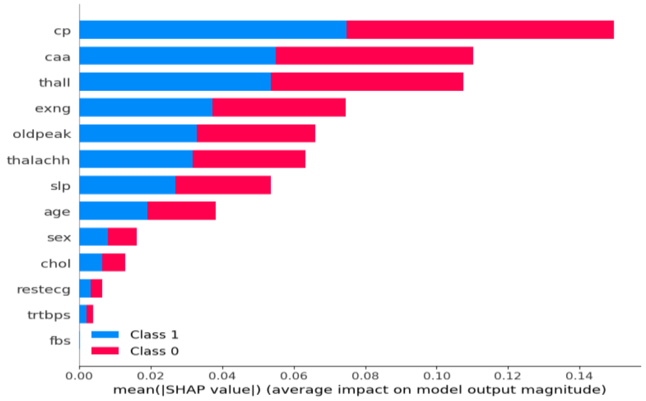


Fig. 7: Summary Plot of SHAP values

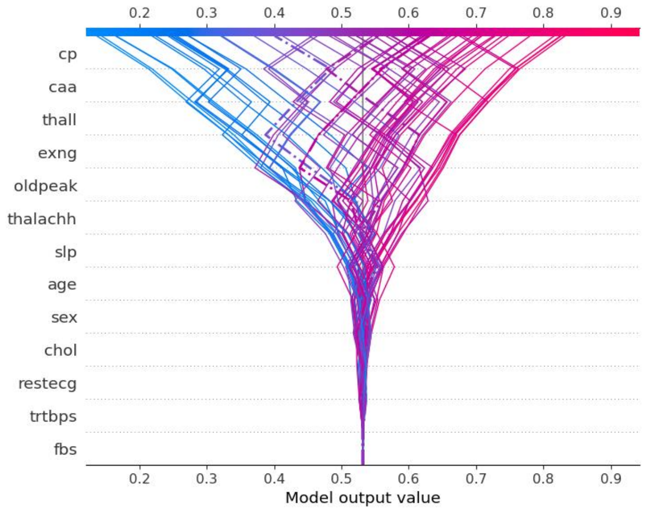


Fig. 8: Decision Plot of SHAP values

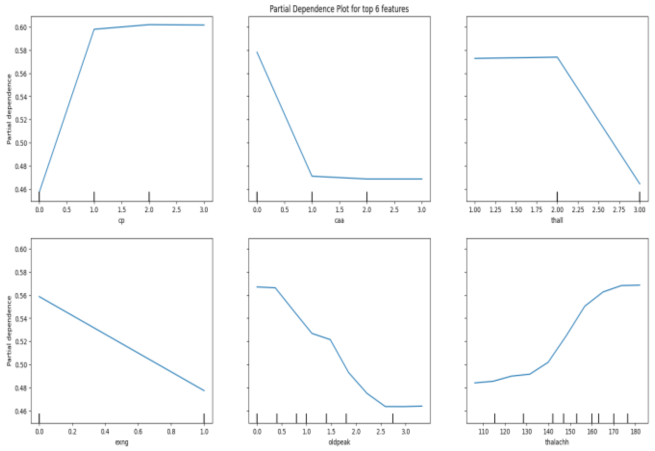


Fig. 9: Partial Dependence Plots of top 6 features

VI. Conclusion

After a comprehensive performance evaluation of various algorithms, it becomes evident that Random Forests, leveraging decision trees as their base estimator, achieves remarkable accuracy of 95%. Notably, Random Forests excel in managing high-dimensional data landscapes, thanks to their ability to extract the most salient features. This proficiency sets them apart from linear classifiers like Logistic Regression and SVM, which lack the capacity to grasp such intricate nuances. In tandem, Explainable AI (XAI) constitutes a pivotal paradigm in AI development, allowing us to demystify intricate model operations and provide transparent insights into their decision-making. In the context of cardiovascular disease prediction, XAI bridges the gap between data-driven predictions and human interpretability, equipping medical practitioners and researchers with a deeper understanding of model's inner workings. As evidenced through the application of SHAP plots, PDPs, and Decision Plots, XAI not only enhances model's credibility but also fosters domain-specific alignment, promoting informed medical decision-making and advancing our quest for accurate disease diagnosis. Thus, in this evolving landscape of predictive healthcare, the fusion of Random Forests and Explainable AI not only raises the bar for accuracy but also illuminates the path towards a future where medical decisions are both reliable and explainable.

References

[1] Ali, Md.Mamun, Bikash.K.P, Kawsar.A, Francis.M.B, J.M.Quinn, and M.A.Moni, Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Computers in Biology and Medicine, vol. 136, 2021.

[2] Prusty, Sashikanta, Srikanta Patnaik, and Sujit Kumar Dash. Comparative analysis and prediction of coronary heart disease, Indonesian Journal of Electrical Engineering and Computer Science 27.2 (2022): 944-953.

[3] M.Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj, Heart disease prediction using hybrid machine learning model, 6th international conference on inventive computation technologies (ICICT), pp. 1329-1333, 2021.

[4] Nissa, Najmu, S.Jamwal, and S.Mohammad, Heart Disease Prediction using Machine Learning Techniques, Wesleyan Journal of Research 13, vol. 67, 2021.

[5] Nawaz, M.Saqib, B.Shoaib, and M.A.Ashraf, Intelligent cardiovascular disease prediction empowered with gradient descent optimization, Heliyon, vol. 7, pp. 5, 2021.

[6] Manjula.P, Aravind.U.R, Darshan.M.V, Halaswamy.M.H, Hemanth.E, Heart Attack Prediction Using Machine Learning Algorithms, International Journal Of Engineering Research & Technology (Ijert), vol. 10, pp. 11, 2022.

[7] S. Manikandan, Heart attack prediction system, 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 817-820, doi: 10.1109/ICECDS.2017.8389552.