xai final report (1).docx

by Ekansh Sharma

Submission date: 28-Nov-2024 12:19AM (UTC+0530)
Submission ID: 2534037875
File name: xai_final_report_1_.docx (1.14M)
Word count: 3777
Character count: 21293

Heart Disease Prediction Using XAI

Utkarsh Singh[1] & Somik Bansal[2]
Utkarsh Singh[1] (BML Munjal University, Computer Science Department, Gurgaon, Haryana, India) Somik Bansal[2] (BML Munjal University, Computer Science Department, Gurgaon, Haryana, India)

Abstract

Cardiovascular diseases still occupy a significant and growing position on the global cause of death list, making timely and accurate predictive models critical. Neural Networks and Random Forests that give high accuracy suffer from low interpretability as they are black boxes. This study uses the XAI approach, focusing on LIME, to provide both, high performance and interpretability for the model to predict heart diseases. In the present analysis, 1025 records comprising 14 features (Age, Cholesterol levels and Type, Chest Pain, etc.) were used and the model for Heart disease prediction named Random Forest Classifier showed good results. Predictor variables including total cholesterol, and maximum heart rate were established. LIME was used to explain each sample's specific prediction and show the most important features in the decision-making model. This combination of machine learning and XAI improves the accuracy and reliability of the model, and thus its relevance in clinical practice. Subsequent studies can improve the model's performance and extend the scale of interpretability to other models.

Keywords: Heart disease, machine learning, explainable AI, Random forest classifier,L IME(Local Interpretable Model-Agnostic Explanations),

1 Introduction

A popular concept in machine learning nowadays, the Explainable Artificial Intelligence (XAI) solves an important problem – the absence of interpretability in models considered as "black boxes." The two main XAI libraries that are used successfully to explain black box machine learning predictions are LIME and SHAP. In this report, we provide a detailed description of end-to-end machine learning for the prediction of the presence of heart disease in patients. To explain these predictions, we use both the LIME and SHAP libraries and unveil the feature that contributes to decision-making. LIME produces explanations by fitting them with simpler models for the instance in question, and SHAP gives importance weights to features for their contribution to a prediction; the two are local and global interpretability. Thus, in this work, we envision the goal to reveal the importance of XAI in increasing machine interpretability and creating the foundation for proper and responsible AI applications in practice.

2 Literature review

The enumerated literature review outlines the increased concern with the yoga poses, classification, correction, and pose estimation fostered by the development in deep learning, HPE, and Explainable AI methods. By integrating Convolutional Neural Networks (CNNs), transfer learning, and/or pose estimation techniques, one of the attempts has been made to overcome hurdles in classifying and correcting yoga poses while considering the complexity in purposes, variations due to human body structure, and the need for immediate feedbacks. This research synthesizes findings from multiple field studies of particular interest, exploring numerous aspects of classification and correction of yoga poses, presenting new solutions based on deep learning models, and explaining ability methods, indicating a collective intention to improve the understanding and applications of machine learning models for yoga poses analysis.

Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and XAI Technique using LIME

This paper discusses how ML and the related sub-corpus of XAI can assist in the early diagnosis of heart attacks, which are a major worldwide cause of death. The authors compared several of the mentioned ML classification methods, namely AdaBoost, Random Forest, Gradient Boosting, and LGBM for the heart attack cases based on clinical data. From these, the LGBM model reached the highest training accuracy which is 99.33 % in this model. The study also pointed that out simple ML technologies that can be categorized as ‘black boxes’ even though they provide very high accuracy rates limit one’s understanding of their functioning. In this regard, the authors used XAI tools: LIME to obtain interpretable representations of found decisions. LIME found the features “kcm” (the medical mark), “troponin” (a protein that is released in the presence of heart muscle damage), and several others as dominant predictors of heart attack. That is to recommend that the considerations of employing XAI in operation with ML models improve not only the forecast’s precision but, more notably, the desperately required transparency necessary for any clinical prescription. In conclusion, the paper suggests that deep learning methods be integrated in order to enhance the accuracy of predictions.

Heart Attack Prediction using Machine Learning and XAI

This work focuses on the use of the ML and XAI methodologies to help predict heart attacks. Many people worldwide lose their lives due to various cardiovascular diseases, though heart attack is one of the most fatal conditions whose early diagnosis can save an individual’s life. The author also incorporates other Machine Learning algorithms such as XGBoost, Logistic Regression, Stochastic Gradient Descent, Support Vector Classifier, K-Neighbors Classifier, and Naive Bayes to forecast heart attacks using a public dataset. Comparing the models of prediction, this study observed that XGBoost was more effective than other models and therefore was the most effective in the prediction of heart attack in the study. Furthermore, the study also involved examining XAI approaches; SHAP, and LIME for explaining model predictions. The described XAI methods allow us to determine which features affect the model’s decision-making most significantly and provide both local and global explanations. This aspect of the research addresses a key issue in healthcare: the requirement for better prediction and models and models that are easily understood by clinicians. The paper concludes on the benefit that may accrue from applying ML with XAI in the future to enhance prediction and management of heart diseases that could inform clinical decisions.

Why Model Why? Assessing the Strengths and Limitations of LIME

In this paper, LIME – Local Interpretable Model-Agnostic Explanations, the widely adopted tool for explainable AI is assessed with a critical lens as a solution for explaining machine learning models in various high-stakes applications including healthcare, finance, and autonomous cars. B. The authors discuss LIME as a tool for improving the interpretability of tabular ML models and use it in a series of experiments with the selected approaches of state–of–the–art machine learning algorithms on the tabular data set. The paper stresses that, as an extensive instrument for providing local explanation, LIME has certain drawbacks, especially if applied to models whereby the input data is in tabular form, not image or text. It is revealed that LIME also can be very sensitive to the sampling process it uses to extract samples and generate explanations, which sometimes can produce very unstable or inconsistent results. However, there should not be lost sight of the fact that LIME is still a valuable tool for enhancing the concepts of model interpretability. The same study also comprises a usability assessment wherein, Eskridge and colleagues LIME introduced outputs to participants who had no prior knowledge of this tool. We find support for the fact that LIME can improve understanding of model predictions but it should be used and interpreted with caution. The authors finally outline changes that can be made to LIME and also propose directions for future studies with other techniques such as SHAP and MAPLE which might provide a more comprehensive input on the workings of ML models.

A Study of LIME and SHAP Model Explainers for Autonomous Disease Prediction

XAI is applied to the recommendation of machine learning models because understanding the output of the models is crucial in healthcare. While offering high accuracy, conventional ML models are more ‘ opaque’ controlling the process as nearing orns and not giving details on the reasoning behind their decisions. LIME (Local Interpretable Model-Agnostic Explanations) introduced by Ribeiro, Singh, and Guestrin in 2016 works at an instance level by explaining the localness of features that influence decisions of the black Box model. In this study, deploying LIME in the prediction of heart disease is discussed with a focus on LIME’s ability to improve model interpretability. That is why by enhancing the understanding of how the ML models work, XAI can foster trust and ethical application in healthcare to enhance clinical decision-making support.

"Table 1: Overview of the Literature Review"

Ref No.	Title	Techniques used	Pros	Cons	Limitations	Dataset used
1	A Study of LIME and SHAP Model Explainers for Autonomous Disease Prediction(**BASE PAPER**)	**LIME** and **SHAP** were applied to a Naive Bayes classifier model to enhance interpretability.	Enhances transparency, trust, and flexibility in AI-driven healthcare predictions.	High computation, limited global insight, and complex interpretation for non-experts.	Limited scalability, accuracy, and consistent interpretability with complex models.	Diabetes prediction Heart disease prediction Breast cancer prediction
2	Unveiling Key Predictors for Early Heart Attack Detection using ML and XAI (LIME)	AdaBoost, Random Forest, Gradient Boosting, LightGBM, LIME	High accuracy (LGBM 99.33%), identifies key predictors like Kim and troponin	AdaBoost performed poorly compared to others	Limited generalizability due to dataset size (Heart Attack dataset with 8 features)	Heart Attack dataset with 1319 samples
3	Heart Attack Prediction using Machine Learning and XAI (SHAP, LIME)	XGBoost, Logistic Regression, Stochastic Gradient Descent, SVM, KNN, Naive Bayes, LIME, SHAP	XGBoost showed the best accuracy, feature importance visualized using SHAP and LIME	High complexity of models, SHAP, and LIME may be difficult to implement for non-experts	Relies on black-box models and complex XAI methods	UCI Heart Disease Dataset
4	Why Model Why? Assessing the strengths and limitations of LIME	LIME applied on Decision Trees, Random Forest, Logistic Regression, XGBoost	LIME provides local explanations for model predictions, enhances model interpretability	Limited to local explanations, lacks global model behavior insights	Limited effectiveness when used for models trained on non-image or complex, high-dimensional data	Various tabular datasets, e.g., the Australian Rain dataset

3 Methodology

3.1 Dataset Collection and Importing

The dataset was obtained from a publicly available source and uploaded for analysis. It included various features relevant to predicting heart disease, such as age, gender, blood pressure, cholesterol levels, and biochemical markers. The data was then imported into the environment for further processing.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
996	56	0	0	134	409	0	0	150	1	1.9	1	2	3	0
274	66	1	0	160	228	0	0	138	0	2.3	2	0	1	1
355	46	0	0	138	243	0	0	152	1	0.0	1	0	2	1
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
702	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1

“Figure 1: Sample of the dataset

3.2 Data Description

❖ **Age**: The age of the patient in years.

❖ **Sex**: The gender of the patient (1 = male, 0 = female).

❖ **CP (Chest Pain Type):** Indicates the type of chest pain experienced by the patient:
0: Typical angina
1: Atypical angina
2: Non-anginal pain
3: Asymptomatic]

❖ **Trestbps (Resting Blood Pressure):** The patient’s resting blood pressure (in mm Hg) on admission to the hospital

❖ **Chol (Cholesterol):** Serum cholesterol in mg/dl.

❖ **FBS (Fasting Blood Sugar):** Whether the patient's fasting blood sugar is > 120 mg/dl (1 = true, 0 = false).

❖ **Restecg (Resting Electrocardiographic Results):** Results of the resting electrocardiogram:
0: Normal
1: Having ST-T wave abnormality (such as T wave inversions or ST elevation or depression of > 0.05 mV)
2: Showing probable or definite left ventricular hypertrophy.

❖ **Thalach (Maximum Heart Rate Achieved):** The maximum heart rate achieved during a stress test.

- ❖ **Exang (Exercise Induced Angina):** Whether the patient experienced angina during exercise (1 = yes, 0 = no).

- ❖ **old peak:** ST depression induced by exercise relative to rest, measuring heart stress.
- ❖ **Slope (Slope of the Peak Exercise ST Segment):** The slope of the peak exercise ST segment:
  0: Upsloping
  1: Flat
  2: Downsloping

- ❖ **CA (Number of Major Vessels Colored by Fluoroscopy):** The number of major vessels (0-3) colored by fluoroscopy (a type of imaging test).

- ❖ **Thal (Thalassemia):** A blood disorder that affects oxygen-carrying proteins:
  1: Fixed defect (no reversible blood flow)
  2: Normal
  3: Reversible defect (reversible blood flow)

- ❖ **Target:** The diagnosis of heart disease (0 = no heart disease, 1 = heart disease).

3.3 Data Cleaning

To ensure data quality, duplicate records were identified and removed. Structural errors in categorical columns were corrected by verifying that values matched predefined valid categories (e.g., specific values for gender, chest pain type, etc.). This step also included filtering out unwanted outliers from numerical columns (e.g., age, systolic blood pressure) using the Interquartile Range (IQR) method to improve data consistency.

- Removing Duplicate Values: 723 duplicate values are removed
- Fixing Structural Errors
- Filtering Unwanted Outliers
- Handling Missing Data

```
Missing values before handling:
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
Missing values after handling:
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

- Validation and Quality Assurance

3.4 Loading Cleaned Dataset

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
35	65	0	2	160	360	0	0	151	0	0.8	2	0	2	1
195	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
234	51	0	0	130	305	0	1	142	1	1.2	1	0	3	0
10	43	0	0	132	341	1	0	136	1	3.0	1	0	3	0
227	58	1	0	146	218	0	1	105	0	2.0	1	1	3	0

3.5 Data Visualization

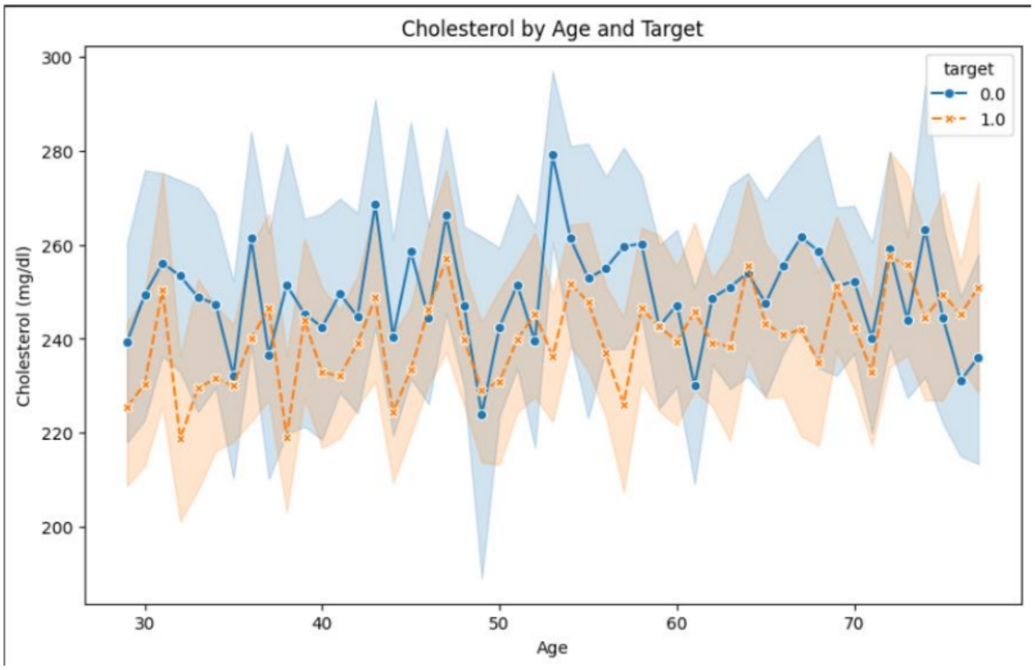


Fig 2: Cholesterol by Age and Target

3.6 EDA (Exploratory Data Analysis)

- Distribution of variables

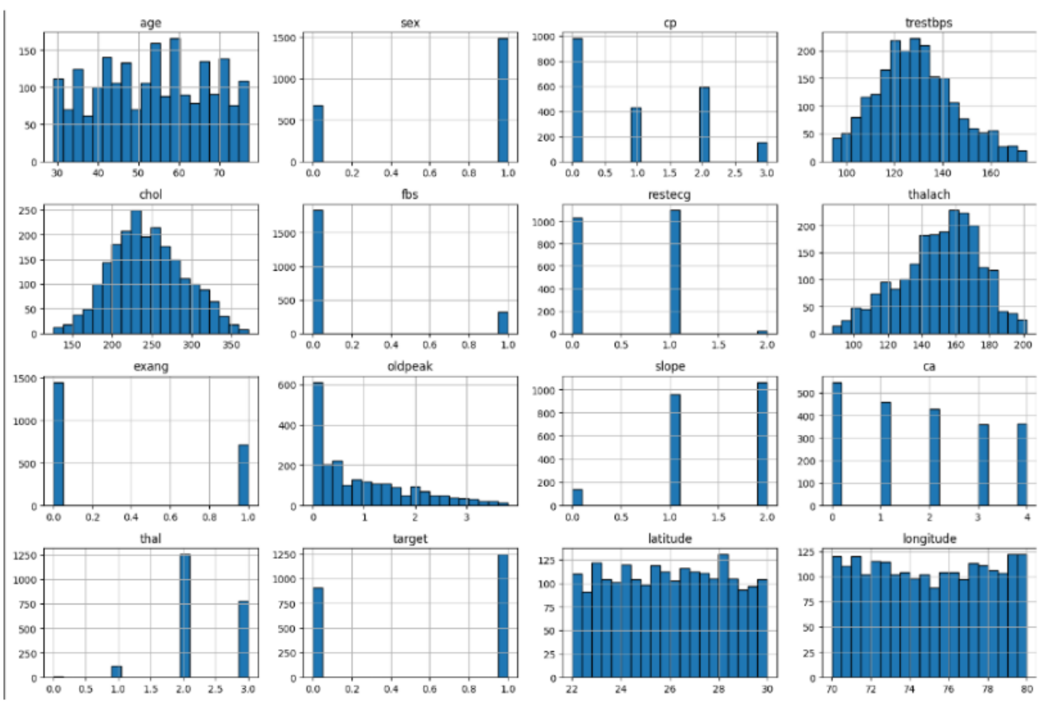
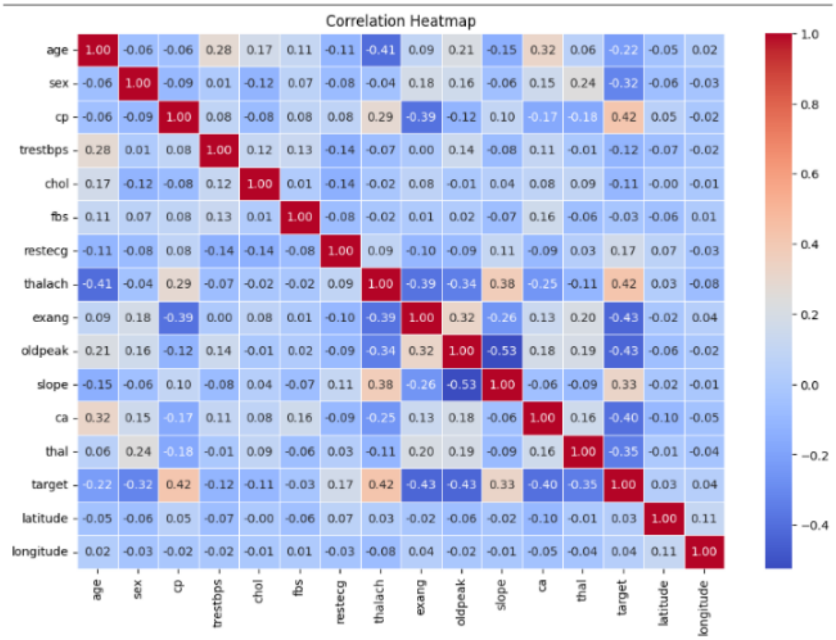


Fig 3 :

- Correlation heat map



- Class Imbalance

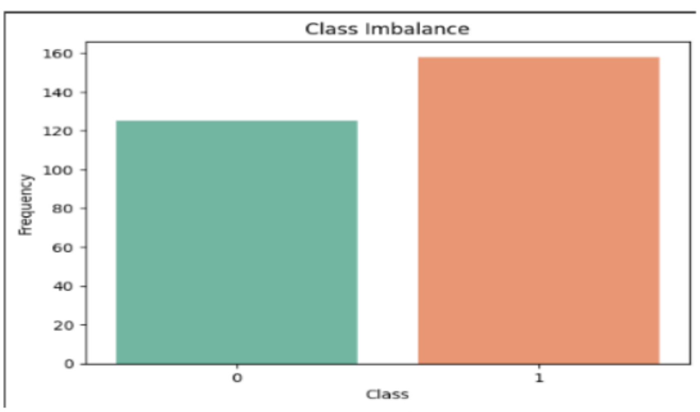


Fig 5

3.7 Model building

- **Random Forest Classifier:** Prediction of heart disease can be done using a Random Forest Classifier by delving into the factors of patients such as age, blood pressure, and cholesterol and coming up with a classification of risk. It utilizes an ensemble of decision trees to enhance its precision and resembles less chance of overtraining thus beneficial in medicinal data sets.

- **Logistic Regression:** The application of Logistic Regression when predicting heart disease involves modeling the probability of such diseases based on probabilities of the feature variables such as age, cholesterol, and blood pressure. This one is perfect for binary classification problems and their application allows us to measure the significance of features on the risk.

- **Decision Tree Classifier:** A Decision Tree Classifier is used to predict heart disease; the data of patients is divided into branches based on some characteristics, such as age, heart rate, etc., related to the risk level. It is easy to implement, easy to interpret, and can be used to model nonlinear relationships in the medical data.

- **XGBoost Classifier:** XGBoost is an Extreme Gradient Boosting Classifier used to predict heart disease; it follows a gradient boosting of mixed model ensemble Type I- decision trees. It brings continuous enhancement, to the forecast's precision by minimizing a loss function and fixing mistakes made by prior trees. The algorithm is efficient, works well with missing values scales well with large datasets, and handles complex data types well. They can model linear and non-linear patterns that exist in medical data and therefore are suitable to be used in the modeling of heart disease.

3.8 **LIME** (Local Interpretable Model-agnostic Explanations)

In our project, LIME (Local Interpretable Model-Agnostic Explanations) was employed to enhance the interpretability of the XGBoost classifier used to predict heart disease. LIME provides a way to explain individual predictions made by any machine learning model, making it an ideal tool to understand complex models.

Key Features of LIME in Our Project:

- Model-Agnostic: LIME works with any machine learning model, regardless of its complexity, allowing us to use it with our XGBoost classifier.

- Local Interpretability: LIME generates explanations for specific predictions by creating a simpler, linear model that approximates the behavior of the more complex XGBoost model for each instance.

- Feature Importance: LIME identifies which features (e.g., age, cholesterol, exercise-induced angina) had the most significant impact on a particular prediction, helping us understand what drives the model's decision.

Application in Our Heart Disease Prediction:

From the above results and using LIME, it was possible to understand how particular predictions of heart disease risk are made. For instance, it assisted in determining if 'Avgsysbp' such as cholesterol levels, 'maxhr' or thalach (maximum heart rate achieved), or 'stdelas' including restecg (electrocardiographic results) were most influential in the positive prediction for heart diseases.

This was a major virtue that helped to lend credibility to the model since in healthcare, besides being able to make the decision, it is equally important to determine why the decision was made. Thus, with the help of LIME, we fixed the problem of lack of trust in the final decision, as all processes in the model were made transparent.

3.8 **SHAP** (Shapley Addictive exPlanations)

As explained in our project, SHAP was used to improve the explainability of the XGBoost classifier applied to the task of heart disease classification. SHAP provides importance value for each feature in respect of a certain prediction which explains how each feature plays its role in model output.

Key Features of SHAP in Our Project:

- Model-Specific and Model-Agnostic: SHAP can be tailored for specific models like XGBoost using TreeExplainer while also offering model-agnostic capabilities, making it a versatile tool for interpretability.
- Global and Local Interpretability: SHAP provides:
    - **Global insight**s: By summarizing feature importance across all predictions (e.g., summary plots).
    - **Local explanations**: By detailing the contribution of each feature for individual predictions (e.g., force plots).
- **Feature Importance and Interaction**: SHAP not only ranks feature importance (e.g., cholesterol, age, exercise-induced angina) but also explores feature interactions, such as the relationship between thalach (maximum heart rate achieved) and ca (number of major vessels colored by fluoroscopy).

Application in Our Heart Disease Prediction:

Using SHAP, we were able to interpret both global and individual predictions made by the XGBoost classifier:

Global Explanations:

Summary plots revealed that features like thalach, chol, and oldpeak were the most influential in predicting heart disease risk.

Local Explanations:

- Force plots illustrated how specific features influenced the prediction for individual patients. For example, a high thalach value might reduce the risk score, while a high oldpeak value could increase it.
- This ability to explain predictions helped us validate the model's decisions, fostering trust and reliability, especially critical in healthcare contexts where the "why" behind predictions is as important as the predictions themselves.

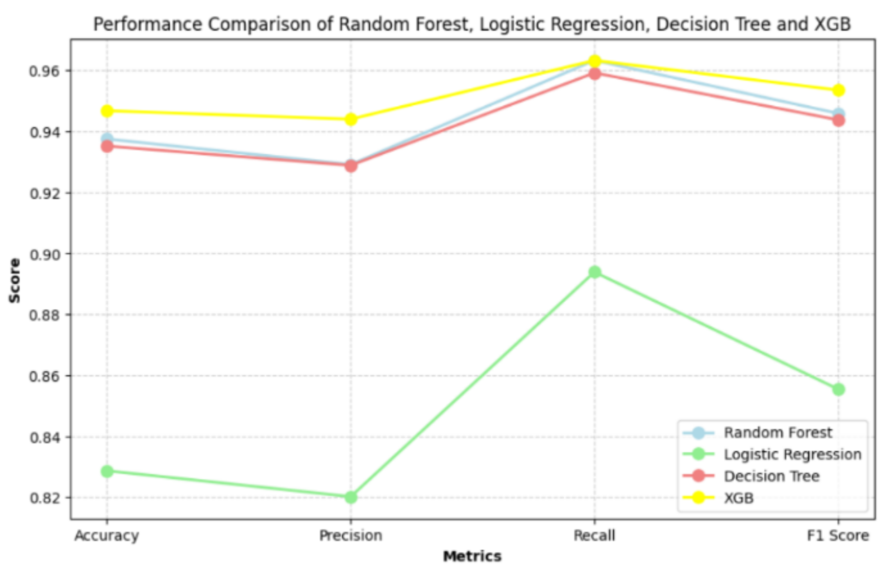
4. Results & Discussion

From the above, the XGBoost model finished with the best performance, and the results in terms of accuracy, precision, recall, and F1-score were 94.68%, 94.40%, 96.33%, and 95.35% respectively. respectively. Random forest was next with an accuracy of 93.75%, and the precision, recall, and F1-score were 92.91%, 96.33%, and 94.59% respectively. Similar to the Decision Tree classifier where it achieved a high accuracy rate of 93.52% its precision, recall, F1- score stood at 92.89%, 95.92% & 94.38% respectively. However, Logistic Regression gave relatively low results and it has got an accuracy of 82.87%, precision of 82.02%, recall of 89.39%, and F1 score of 85.55%.

Model	Accuracy	F1 Score	Precision	Recall
XGBboost	0.9468	0.9535	0.9440	0.9633
Random Forest	0.9375	0.9459	0.9291	0.9633
Decision Tree	0.9352	0.9438	0.9289	0.9592
Logistic Regression	0.8287	0.8555	0.8202	0.8939

Table 2 : Model Differentiation

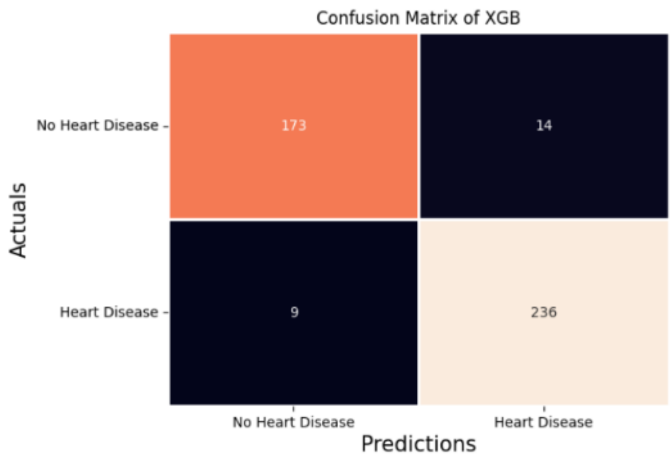
Among all the algorithms used, XGBoost has proved to be the best all-rounded by being slightly more accurate than Random Forest and decision tree classifiers. To validate the models, cross-validation was used thus showing disparities in the models' performance in different datasets. Among the mentioned algorithms, XGBoost was on top of all of them followed by Random Forest, Decision Tree, and Logistic Regression in that order.



"Figure 6: Comparison graph

The XGBoost Classifier's confusion matrix highlights key performance outcomes: That true positive (TP) number of cases of the virus that were correctly identified; true negatives (TN), the number of people who tested negative for the virus and are indeed, negative; false positives (FP), the number of negative individuals who tested positive for the virus; False negatives (FN), the number of individuals with the virus who tested negative. Recall is particularly relevant in predicting heart diseases with an emphasis on accurately identifying positive cases. High recall is important in medical diagnosis especially where the costs of missing the diagnosis perhaps due to a low recall are likely to be severe for the patient. In particular, the recall of XGBoost was 96.33%, which means that the drug checked about 2% of actual cases of heart disease during searches. This reduces potential and modifies missed diagnoses, thereby promoting patients' safety and efficacy. Moreover, XGBoost keeps a good proportion of the recall rate and the precision rate, eliminating misdiagnosis of both types of losses and gains for more credible and accurate outcomes.

40



"Figure 7: Confusion Matrix

- **LIME EXPLANATION FOR XgBoost MODEL**

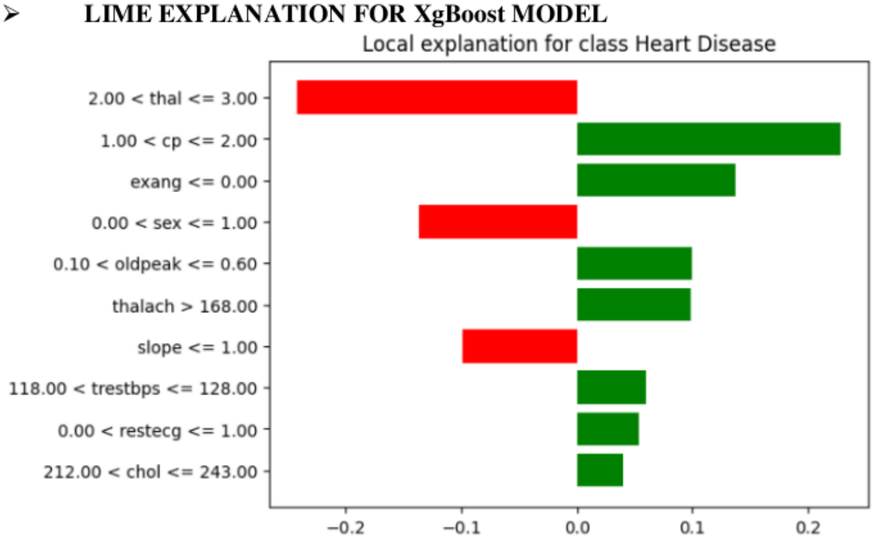


Fig 8 :Local explanation for class Heart Disease

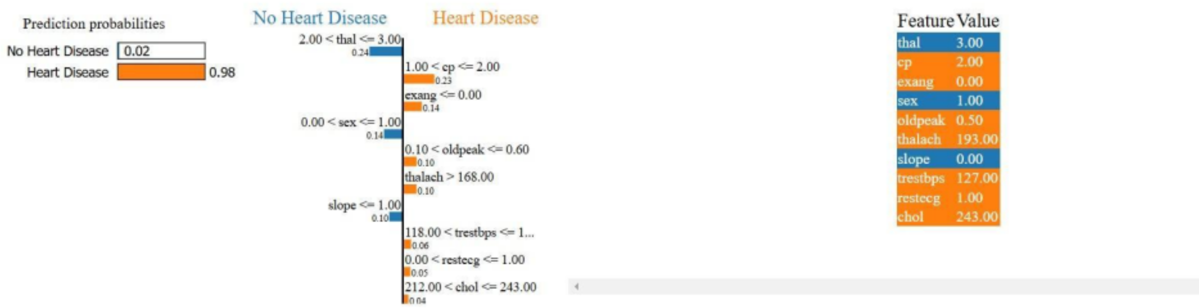


FIG 9: No Heart Disease (blue color) and Heart Disease (orange color)

Features pushing the prediction towards Heart Disease (Orange color):

- **cp (Chest Pain Type)**: A value of 2.00 (non-anginal pain) moderately pushes the prediction towards heart disease, as certain chest pain types are associated with cardiac issues.
- **exang (Exercise-Induced Angina)**: A value of 0.00 (no exercise-induced angina) slightly contributes to heart disease, as the absence of angina during exercise could still indicate underlying cardiac risks.
- **oldpeak:** An ST depression value of 0.50, caused by exercise, contributes to the prediction of heart disease, as even mild ST depression indicates stress on the heart.
- **thalach (Maximum Heart Rate Achieved):** A high value of 193.00 slightly contributes to heart disease, as extremely high heart rates may indicate abnormal stress responses or cardiac conditions.
- **trestbps (Resting Blood Pressure):** A resting blood pressure value of 127.00 contributes slightly towards heart disease. While in a near-normal range, it can still indicate mild cardiac stress.
- **restecg (Resting Electrocardiographic Results)**: A value of 1.00 (ST-T wave abnormality) moderately pushes the prediction towards heart disease, as it often reflects irregularities in heart activity.
- **chol (Cholesterol):** A cholesterol level of 243.00 contributes to the prediction of heart disease, as high cholesterol is a known risk factor for cardiovascular conditions.

➢ SHAP(SHaply Addictive exPlanations)

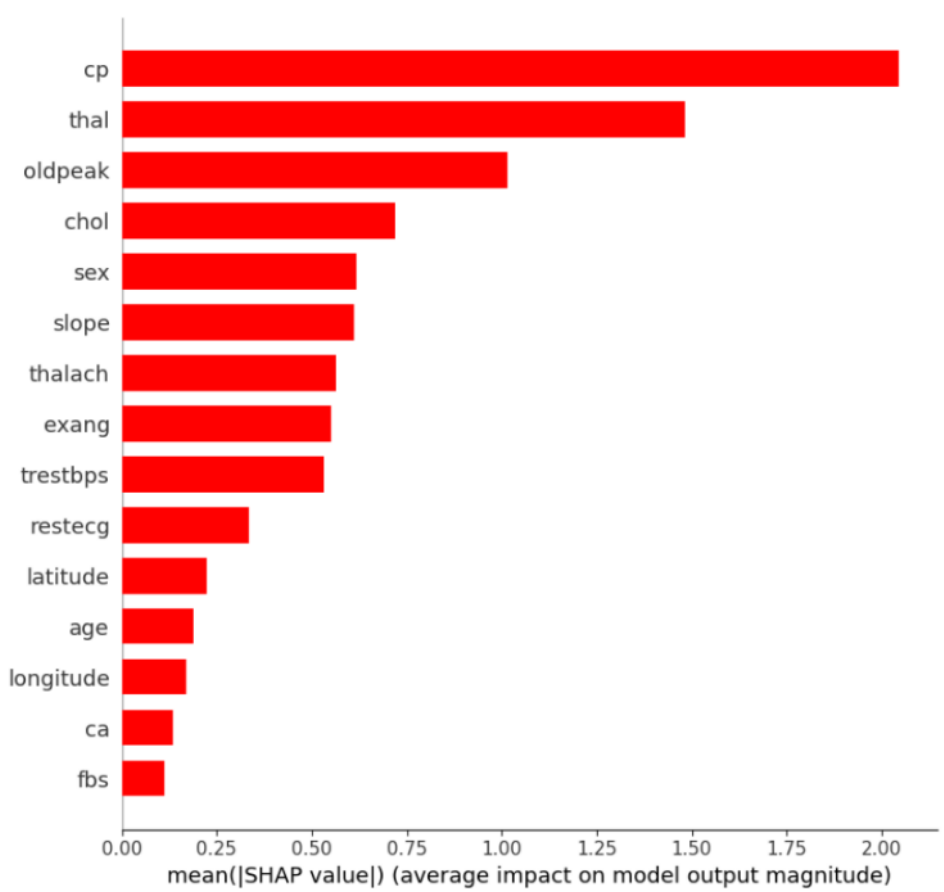
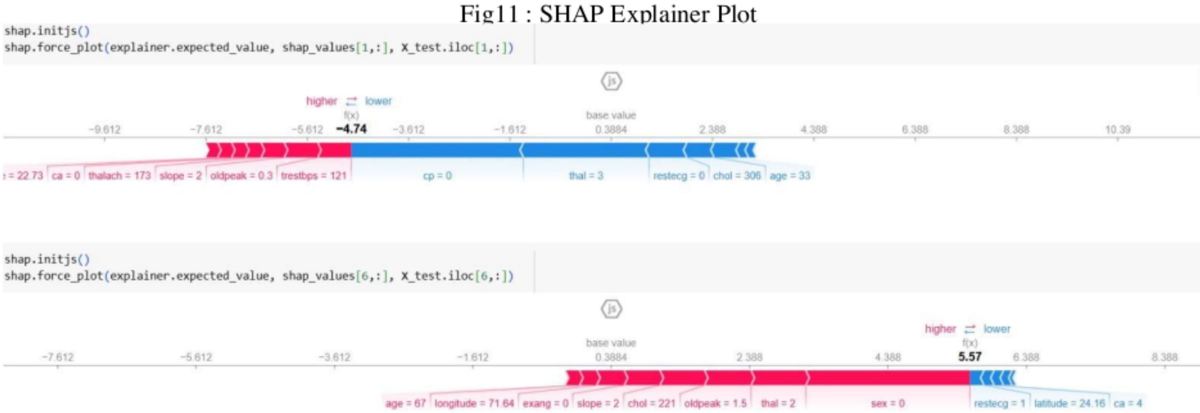
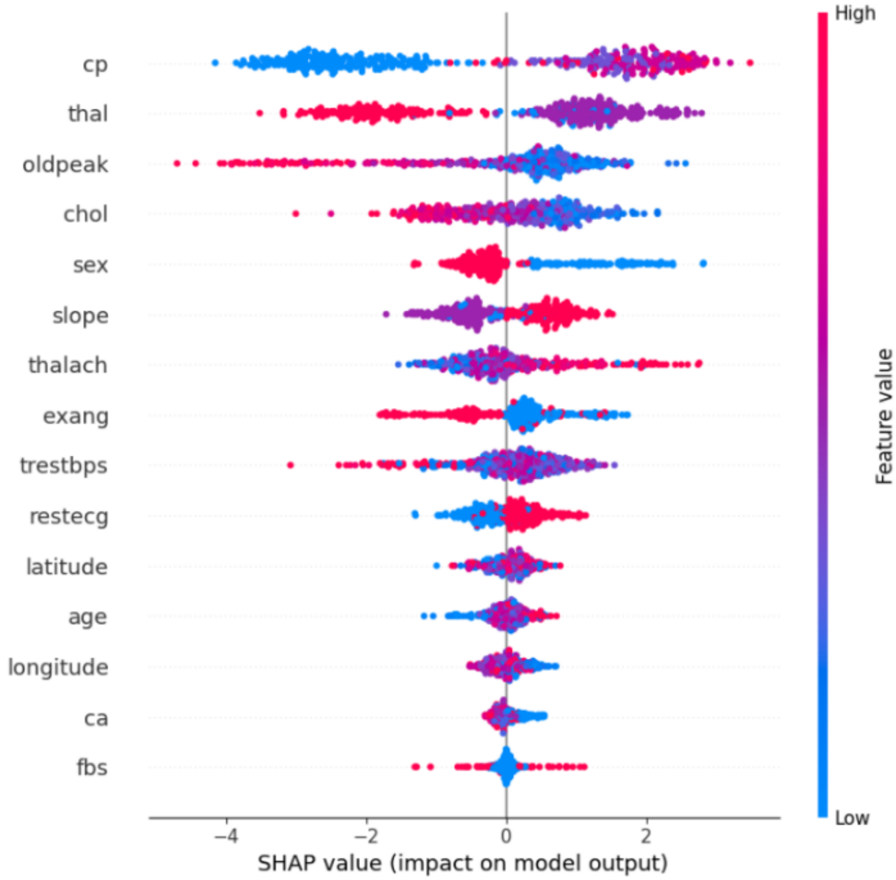


FIG 10

Fig11 : SHAP Explainer Plot

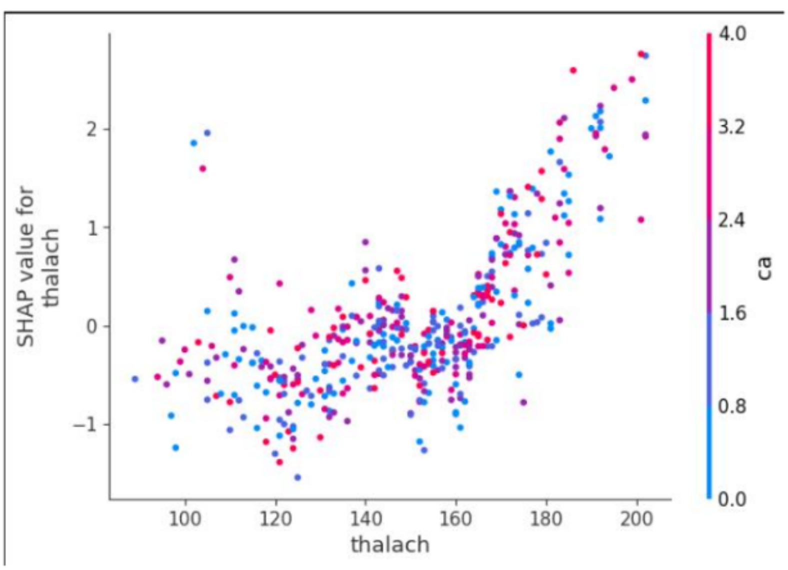




High
cp
thal
oldpeak
chol
sex
slope
thalach
exang
trestbps
restecg
latitude
age
longitude
ca
fbs
Feature value
Low
−4
−2
0
2
SHAP value (impact on model output)

➤

FIG 12



SHAP value for thalach
2
1
0
−1
100
120
140
160
180
200
thalach
4.0
3.2
2.4
1.6
0.8
0.0
ca

FIG 13

5 Conclusion

Because medical diagnosis involves the prediction of diseases such as heart diseases, false negatives, that is, actual patients misclassified as healthy, are disastrous. Thus, the target is to select the model with high recall which describes the model's ability to predict true positive cases. The measures of precision, or accuracy, and recall as well as overall accuracy are used in evaluating the performance of machine learning models Other factors used in the evaluation of the models include; successful reporting of the disease by the models as measured by recall takes precedence with sick patients always being identified.

When using Logistic Regression, Decision Tree, Random Forest, and XGBoost classifiers it was found that the XGBoost was providing better results than the rest every time. Not only that, this model had the highest recall compared to the other models it also has higher precision, F1-score, and accuracy. Most importantly, the false negative number of the XGBoost model is the smallest, which raises the potential for using it in the diagnosis of potential heart disease and early treatment for patients. Recall and precision of Random Forest and Decision Tree classifiers were also high and the Random Forest approached that of XGBoost. The logistic regression model, as well, had average results and ranked lower than most models in most of the assessments.

Therefore it is concluded that the XGBoost model is more efficient and accurate in predicting heart disease, which is the best model to select for the applications where correct identification of abnormal patients is necessary. This analysis underscores the importance of the use of multiple evaluation metrics while training and selecting a machine learning model, to arrive at a model that can give both highly accurate and clinically useful predictions.

xai final report (1).docx

ORIGINALITY REPORT

14%	8%	10%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Arab Open University Student Paper	2%
2	Submitted to Sheffield Hallam University Student Paper	1%
3	Submitted to Coventry University Student Paper	1%
4	Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	1%
5	www.mdpi.com Internet Source	1%
6	www2.mdpi.com Internet Source	1%
7	Submitted to BML Munjal University Student Paper	<1%
8	www.researchgate.net Internet Source	<1%

9	"Proceedings of International Conference on Recent Innovations in Computing", Springer Science and Business Media LLC, 2024 Publication	<1 %
10	Miguel A. Labrador, Oscar D. Lara Yejas. "Human Activity Recognition - Using Wearable Sensors and Smartphones", Chapman and Hall/CRC, 2019 Publication	<1 %
11	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1 %
12	Submitted to The University of the West of Scotland Student Paper	<1 %
13	Submitted to German University of Technology in Oman Student Paper	<1 %
14	Submitted to University of North Texas Student Paper	<1 %
15	Geeta Rana, R. K. Tailor, Vugar Abdullayev. "Data-Centric AI Solutions and Emerging Technologies in the Healthcare Ecosystem", CRC Press, 2023 Publication	<1 %
16	Submitted to CSU, San Jose State University Student Paper	<1 %

17	K. Karthick, S. Ravivarman, R. Priyanka. "Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation", World Electric Vehicle Journal, 2024 Publication	<1 %
18	Submitted to Pathfinder Enterprises Student Paper	<1 %
19	Submitted to University of Portsmouth Student Paper	<1 %
20	linnk.ai Internet Source	<1 %
21	K Hemachandran, Debdutta Choudhury, Raul Villamarin Rodriguez, Jorge A. Wise, T Revathi. "Predictive Analytics and Generative AI for Data-Driven Marketing Strategies", CRC Press, 2024 Publication	<1 %
22	R. Indrakumari, T. Ganesh Kumar, D. Murugan, P.C. Sherimon. "Deep Learning in Medical Image Analysis - Recent Advances and Future Trends", CRC Press, 2024 Publication	<1 %
23	fidelity.nusaputra.ac.id Internet Source	<1 %
24	scholarworks.rit.edu Internet Source	<1 %

25	www.frontiersin.org Internet Source	<1 %
26	Nidhi Gupta. "Applications of Optimization and Machine Learning in Image Processing and IoT", CRC Press, 2023 Publication	<1 %
27	github.com Internet Source	<1 %
28	ijetms.in Internet Source	<1 %
29	towardsdatascience.com Internet Source	<1 %
30	www.journal-aquaticscience.com Internet Source	<1 %
31	Hang Wu, May D. Wang. "Infer Cause of Death for Population Health Using Convolutional Neural Network", Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB '17, 2017 Publication	<1 %
32	Norah Saleh Alghamdi, Mohammed Zakariah, Achyut Shankar, Wattana Viriyasitavat. "Heart disease prediction using autoencoder and DenseNet architecture", Egyptian Informatics Journal, 2024	<1 %

Publication

Exclude quotes	On	Exclude matches	< 6 words
Exclude bibliography	On		