Early Cardio Vascular Disease Detection using Machine Learning and Explainable AI

Interim Report

by

Bhargava Dandu

Master of Science in Data Analytics

Dundalk Institute of Technology

Supervisor

Dr Zohaib Ijaz

June 2025

1 | Background

1.1 Introduction

As per statistics of World Health Organization (WHO), Cardiovascular diseases (CVDs) cause 32% of global deaths worldwide and 28.1% deaths in India. The severity of the patient depends on the number of years he is taking to recover after getting diagnosed. Studies show 85% of the patients recover within one year of treatment, but only 55% of the patients recover after two years of treatment, and only 33% of patients revover within 5 years of treatment [11]. This shows the criticality of early detection of cardiovascular diseases.

1.2 Research Questions

- How can we apply effective Machine Learning(ML) techniques for early detection of heart diseases on publicly available datasets?
- In what ways do XAI techniques like SHAP and LIME help to improve the interpretability of the predictions made by the ML model?
- How does enhancing the transparency of the models improves the trust, confidence in the decision making of the healthcare professionals?

1.3 Problem Definition

Cardiovascular diseases (CVDs) are one of the major causes of the deaths worldwide. Using Machine learning models many researchers developed complex models to identity early signs of heart diseases in patients. But, the black box models are very complex to interpret and we cannot the reason behind the prediction. In such situations, even though we develop a model with high accuracy, medical industry experts would lack trust on it, because the false positives and false negatives are so valuable in medical industry. So, to overcome this problem statement, we are trying to incorporate Explainable AI techniques (XAI) to make the models transparent. If the models are transparent and we are able to interpret their predictions then medical industry experts would develop trust, confidence in their decision making by using our models.

1.4 ML-Based Solution

Artificial Intelligence (AI) and Machine Learning (ML) technologies are reshaping the medical industry every day. So, in this research, we aim to use Machine Learning techniques and Explainable AI (XAI) techniques for early detection of cardiovascular diseases. We implement the ML techniques on the dataset obtained from UCI machine learning repository and further apply XAI techniques for further explanations of model's predictions. These techniques clearly outline which feature is influencing the model's prediction which makes the decision making process more transparent.

1.5 Core Technologies Used

In our research, we are planning to incorporate widely used XAI frameworks like SHAP (SHapeley Additive exPlanations) and LIME (Local Interpretable Model-agnositic Explanations). These tools helps us to interpret the model's classification by highlighting the individual feature contributions towards the model outcome. This will provide transparency in our model's classifications and help us to gain trust from medical industry experts [10]. By incorporating XAI techniques, we can see transparency in results. This will help the medical industry experts to understand which parameters contribute more for final classification.

1.6 Proposed Hypotheses

In this research we are aiming to combine predictive accuracy using ML techniques and clear interpretability using XAI techniques which enhances the trust in the medical professionals.

2 | Literature Review

Data Preprocessing and Ensemble Learning techniques

Authors in [1] has used good data preprocessing techniques before applying ML models. The researcher has incorporated ensemble techniques like Bagging, Boosting to get the final predictions and narrowed down the final predictions using voting technique. In their research, they have used standard heart disease dataset to make predictions whereas in real-time the heart disease data may contain some uncertainties .

In the same way authors in [13], have implemented ML and DL models on standard dataset with only 303 instances. For industry advancement we must try to integrate real-time dynamic data and train our model to make better classifications .

Authors in [9] have focused on building a pipeline to integrate live patients records as an input and classify the patients. They have used Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) for data pre-processing and extracted important features. In this research they have used novel techniques to handle the missing values and they have used covariance based multivariate outlier detection procedure to detect the outliers and remove them.

In our research, As a novel approach, we aim to train our model using the same dataset, but we are trying to incorporate XAI techniques like SHAP and LIME to explain our model predictions. This results in better understanding for both the doctors and researchers, which variables play an important role in classifying the presence of heart disease.

Hyperparameter tuning and Feature Engineering

Authors in [3], implemented GridSearch CV technique used for hyperparameter tuning. They have used five fold cross validation technique to train their model and employed soft voting ensemble method to increase model predictions. Though they have used hyperparameter tuning, they did not focus much in feature engineering techniques.

Similarly, authors in [2], has focused on using feature selection techniques to extract appropriate features for model training. But, the feature selection techniques used in this model are static and they may lead to increase biasness in the predictions.

So, in our research we aim to use feature engineering techniques to create new features and try to use XAI techniques to know which feature contributed more for model predictions. For eg: if we combine a person age and LDL cholesterol then we can create a new feature and focus on estimating the risk of early heart attacks of the patients.

Explainable AI (XAI) for Model Transparency

Authors in [6], tried to classify the Hate Speech(HS) by using machine learning and deep learning models and evaluated the models performance. They have used Google Zigsaw and HateXplain to train different machine learning and deep learning models respectively. Out of the existing machine learning models, LSTM model has shown the best performance when trained on Google Zigsaw dataset with an impressive accuracy and recall score. To achive this result, they have used Count Vectorization and TF-IDF techniques for feature extraction. After pre-processing they have implemented decison tree, KNN, random forest, multinomial naive Bayes, LR and LSTM models on the google jigsaw dataset. They have implemented BERT + MLP and BERT + ANN models on the HateXSpeech dataset. Further to evaluate the model's predictions they have implemented XAI techniques like LIME on BERT variants and not other models because Google jigsaw dataset is not annotated like HateXSpeech dataset. By using XAI techniques, they further fine tuned the model for 50 epochs.Though this paper is not related to cardio vascular disease, this paper briefly explained about XAI techniques and how they play major role in interpretting model's predictions. Based on these techniques we can have better understanding of the model's algorithm and individual feature's contribution on the final prediction.

Authors in [5], used the Random Forest machine learning model to predict the mortality of the patient. They major contribution with this research is the feature extraction techniques used and the implementation of XAI techniques like SHAP to explain the reasoning behind the model predictions. For the feature selection, they have used LASSO regression(Least Absolute Shrinkage and Selection Operator) to shrink the less relevant features to 0. Then further they implemented the RFE technique to recurssively eliminate the unwanted features and reduced the dimentionality of the training data. By using these techniques they have achieved good model performance and further implemented XAI techniques to explain the probability of the features which contribute to the model predictions.

Authors in [8], performed the standard feature engineering technique which is Principle Component Heart Failure (PCHF) feature engineering technique and converted the standard features to reduced features. This technique extract the important components from the dataset and they are further implemented into the ML models. By using this technique they have achieved good results and the Decision trees, Random Forest algorithms have outperformed the other ML models (LR, SVM, NB, MLP). From their research we can consider two major gaps, first they did not explain the importance of newly created features which we include in our research. Secondly, the real time heart disease data is very dynamic and can be varying from patient to patient. So, we can propose few feature engineering techniques on multiple sub groups of patients as our future study.

In our research, we aim to create new features, and check the LASSO Regression values of them and proceed to implement our ML model on it. We can also further use the same XAI techniques used in this paper and evaluate the importance of newly created features.

Integration with IoT and Real-Time Systems

Authors in [4], aimed to integrate AI framework with ML and IoT technologies. This approach is highly used in medical industry these days. They have compared the classification models within each and also with/without using the XAI techniques. Out of them, they have realised that models with XAI techniques are more trust worthy, when compared with the one that do not use XAI. They have used ealuation metrics like AUC and ROC curve, sensitivity and specificity to evaluate the model performances. The SVM classifier score high accuracy among all the implemented models. Techniques like feature selection, explainable feature weight initialisation, normalisation and optimisation have been used in this paper along with XAI. We can critically appreciate their approach where they tried to integrate the ML technologies with IoT devices. They have also used XAI technology which acts like a bridge for medical industry and researchers for both of them to better understand how the ML models are classifying patients. They have considered secondary evaluation done by doctors to check if the model has taken right parameters to make the predictions. In this particular research, they have mentioned their limitation that they trained their model with a dataset that contains limited variables. They have suggested to implement same techniques on other complex datasets as their future study. We can bridge this research gap in our thesis by implementing similar techniques in our dataset.

Deep Learning and ECG-Based Predictions

Authors in [7], have applied algorithmic models to analyze ECG signal in an eiffective way. They used 2 deep learning algorithms Mobile net V2 and VGG16 to classify ECG images. To achieve this they have implemented the Deep Learning models on the ECG images dataset. We can critically appreciate their approach of reducing the dimensionality of the dataset. Because Deep Learning models have complex algorithms and we do not know what happens inside those algorithms. So reducing the dimentionality using brightness, contrast, gamma, hue, saturation and central-crop is very brilliant. But, They did not explain the model's predictions. They are focussing on deep learning techniques and hyperparameter tuning techniques to optimise the model performance, but they are not considering which feature contributed more for model prediction. If they can find out that, they can gather a dataset accordingly which reduces the effort in dataset pre-processing. Authors in [12], used 4 different ML models to classify the heart disease of patients. They have taken 80/20 split of the train/test dataset and trained, Naïve Bayes, Random Forest, Decision tree and Logistic regression ML models. They have used F1- Measure, Precision, Accuracy and recall as their evaluation metrics to evaluate the performance of the model. Based on their research, they have classified that Random forest has the highest accuracy among the models. In our research we aim to consider this research as base and work their future work by implementing Explainable AI techniques on ML model to explore reasoning of model's classification.

3 | Exploration of Data and Methods

3.1 Dataset Collection

The data in this dataset is collected by clinical data collection from the patients who were taking angiographic test. Angiography is a test conducted to check arteries of human heart by injecting dye into the blood. After angiography, the researchers have extracted the test results from the Electronic Health records (EHRs). Then they have extracted 14 important features from the overall features of the dataset. Then they have anonymised the extracted data for research purpose. Further they have aggregated the data as per region.

3.2 Dataset Description

We have obtained this licensed dataset from UC Irvine Machine Learning Repository. This dataset is formed by gathering patient information who are undergoing angiogram treatment in Cleaveland, Hungary, Switzerland and Long Beach VA. The raw format of these individual data files contains 76 attributes and 303, 294, 123, and 270 instances for Cleaveland, Hungarian, Switzerland, and Long Beach VA respectively. Out of the 76 attributes, researchers have extracted 14 attributes that are ideal for predicting heart diseases. The target variable is one of the 14 attributes.

Feature	Description
age	age of the patient
gender	sex of the patient
chest_pain	type of chest pain the patient is experiencing, severity from (0-4)
resting_BP	patients BP while resting (mmHg)
serum_cholestrol	The level of cholestrol in patient's blood (mg/dl).
fasting_blood_sugar	patients blood sugar level, 0 being normal and 1 being high.
resting_ecg	result of ECG test, 0 being normal, 1 being ST-T wave abnormality and 2 being left ventricular hypertrophy.
max_heart_rate	the maximum number of heart beats in a minute
exercise_angia	patient feels chest pain during physical activity or not.
old_peak	A measure from ECG test showing the drop in the heart rate during exercise.
slope	describes how heart activity changes during physical activity.
no_of_major_vessels	number of blood vessels present in heart.
thal	Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)
target	Diagnosis of heart disease (0 = no disease, 1 = mild, 2 = severe)

Table 3.1: Features Description

3.3 Preliminary Analysis

After performing the Uni-variate analysis on the features present in our dataset, we got an overview of how data is distributed in the dataset. Here are the key points that outlines the features as follows:

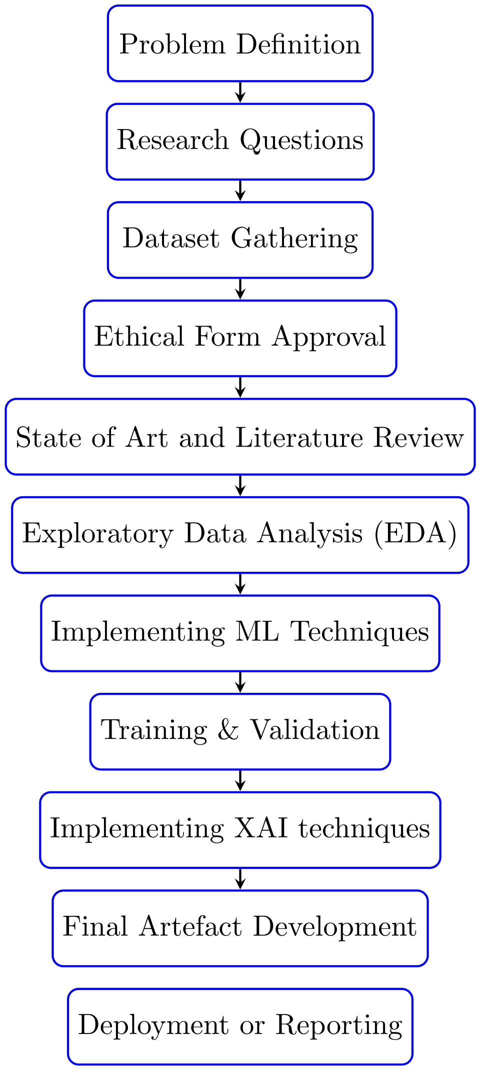
- The dataset is created by extracting ECG results from male patients, the male patients are twice greater than female patients.
- The age of the patients present in the dataset is normally distributed between 30 years and 75 years.
- The patients with chestpain severity 4 are 50% in this dataset.
- serum cholesterol has normal distribution but there are few outliers. Some values in the dataset are containing 0, which is technically not possible. So, we shall fill the serum_cholesterol values using linear regression.
- The fasting blood sugar of majority of the patients is under normal conditions (¡= 120 mg/dl), only 45(approx) patients blood sugar level is higher (¿120mg/dl).
- maximum heart rate of patients is showing normal distribution with a mean around 150- 160 beats per minute.
- Since majority of the patients are involved in physical activity, the patients without exercise angia are 2x times than patients suffering from exercise angina.

- The slope of peak exercise ST segment contains 3 different classes and out of them downsloping class patients are very less in number.

After performing Uni-variate Analysis we came to understand that the patients classes are not equally distributed. So, when we train our model, we must make sure when training our model, we must use stratified cross-validation to split the test/train split.

4 | Proposed Future Analysis

4.1 Artefact Development Lifecycle



5 | Conclusion

This research highlights the importance of requirement for early detection of heart diseases among people. By using Machine Learning and XAI techniques on the publicly available datasets, we were able to identify hidden trends in the heart disease dataset and what are the metabolic markers causing cardiovascular diseases. After performing EDA on the heart disease dataset we came to know the distribution of the classes in the dataset and which features are highly correlated with each other.

Beyond Prediction and accuracy, interpretability is equally important in medical industry so, we are incorporating XAI techniques like SHAP and LIME to explain the predictions of the model.This transparency bridges the gap between model predictions and trust of medical industry experts. This research suggests transparent models are more trustworthy and supports decision making process in health-care industry.

Bibliography

[1] Absar, N., Das, E. K., Shoma, S. N., et al. 2022in , MDPI, 1137

[2] Ahmad, B., Chen, J., & Chen, H. 2025, arXiv preprint arXiv:2503.16577

[3] Chandrasekhar, N., & Peddakrishna, S. 2023, Processes, 11, 1210

[4] Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N., & Alarfaj, F. K. 2022, Electronics, 11, 4086

[5] Lee, H., & Tsoi, P. 2025, arXiv preprint arXiv:2503.21241

[6] Mehta, H., & Passi, K. 2022, Algorithms, 15, 291

[7] Mhamdi, L., Dammak, O., Cottin, F., & Dhaou, I. B. 2022, Biomedicines, 10, 2013

[8] Qadri, A. M., Raza, A., Munir, K., & Almutairi, M. S. 2023, IEEE Access, 11, 56214

[9] Rajendran, R., & Karthi, A. 2022, Expert Systems with Applications, 207, 117882

[10] Rezk, N. G., Alshathri, S., Sayed, A., El-Din Hemdan, E., & El-Behery, H. 2024, Bioengineering, 11, 1016

[11] Sethi, A., Dharmavaram, S., & Somasundaram, S. 2024, in 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), IEEE, 1–6

[12] Sharma, V., Yadav, S., & Gupta, M. 2020, in 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN), IEEE, 177–181