Early Cardio Vascular Disease Detection using Machine Learning and XAI *

Bhargava Dandu
Dundalk Institute of Technology
D00274867@student.dkit.ie

***Abstract*—Cardiovascular diseases are the leading causes of the death worl wide by contributing 32% of deaths world wide. In Majority of the fatal cardiovascular disease cases, the medical industry experts identify the symptoms at the final stage. So, early detection of the cardiovascular disease are crutial. In this research we use Machine Learning (ML) and Explainable AI(XAI) to identify which parameters to be considered to identify the disease in early stages. Using a dataset that we have obtained from UCI machine learning repository, we have implemented Random forest classifier to predict the cardio vascular disease of a patient and achieved a model performance of 62%. We further tuned the model with Ml techniques and created 2 additional models with a better model accuracy with 80% of F1 score for the Gradient boosting classifier.The novelty of this research is how we handled the NULL values and creating features that will help ML models to discover deeper trends in the dataset. The Explainable AI(XAI) techniques are used to understand the decision making of the model and factors influencing the predictions. The future work of this research is refining the classifications based on the newly created features and further classifying the severity of the heart disease of the patient using them.**

Keywords- Cardio Vascular Diseases(CVDs), SHapley Additive Predictions(SHAP), Machine Learning (ML), Eplainable AI (XAI), Local Interpretable Model-agnostic Explanations (LIME), Gradient Boosting, Precision, Recall, F-1 Score,

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 [10].

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.

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 and reliable.

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 [9]. 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. This research report consists of the following sections Section 2: Literature review, Section 3: Exploration of Data and Methods, Section 4: Machine Learning Techniques, Section 5: Explainable AI techniques, Section 6: Evaluation, Section 7: Conclusion and Future work, Section 8: Acknowledgment to our supervisor and followed by References.

2. Literature Review

Authors in [2] 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 [11], 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 [6] 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.

Authors in [7], 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 [12], 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.

Authors in [4], 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 [13], 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.

Authors in [3], 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.

Authors in [5], 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 [1], 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.

TABLE 1. Description of Dataset Features

Feature	Description
age	Age of the patient
gender	Sex of the patient
chest_pain	Type of chest pain (0–4, increasing severity)
resting_BP	Resting blood pressure (mmHg)
serum_cholestrol	Serum cholesterol level (mg/dl)
fasting_blood_sugar	Fasting blood sugar (0 = normal, 1 = high)
resting_ecg	ECG result (0 = normal, 1 = ST-T abnormality, 2 = LV hypertrophy)
max_heart_rate	Maximum heart rate achieved
exercise_angia	Exercise-induced angina (yes/no)
old_peak	ST depression induced by exercise
slope	Slope of the ST segment during exercise
no_of_major_vessels	Number of major blood vessels (0–3)
thal	Thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect)
target	Diagnosis (0 = no disease, 1 = mild, 2 = severe)

3.3. Exploratory Data 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 MLP 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.

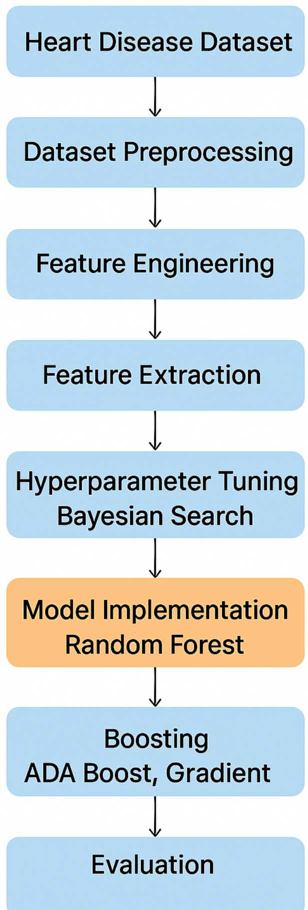

Figure 1. Methodology.

3.4. Data Pre-processing

In this research, one of the novel approach we have taken is the way of how we performed data pre=processing. Since we deal with medical data, every medical record is important. So, we did not delete a single record from the dataset that we have collected. For the columns that have less than 5% of missing records in them, we have replaced them with the median, median or mode values depending on the distribution of the column. For the columns that are containing missing values of more than 30% we have used KNN imputation approach to fill the NULL values.

After handling the NULL values, we found out that, there are few records in serum_cholesterol column that are zero. Technically, serum_cholesterol cannot be zero. So, we used Multi Layer Perceptron Regression model to predict the values that are zero, by figuring out hidden relationships of the dataset. We have used age, resting_bp and the max_heart_rate columns to predict the serum_cholesterol values. Further we replaced the original serum_cholesterol values with the predicted values.

4. Machine Learning Techniques

After pre-processing the dataset, we have now build a model that predicts the presence of heart disease in a patient. We have used random forest model to build a base-line model, because since our dataset has mixed data types, and there were many outliers in this dataset as well. Further now we shall proceed to fine tune this model by using machine learning techniques to improve the model efficiency.

4.1. Hyperparameter Tuning

In this research, we have used bayesian search to check for the right hyperparameters instead of grid search or random search. The reason we have used bayesian search is this technique is computationally efficient when compared to other searching options. Also, provides best hyperparameter values by efficiently searching for them. After performing the hyperparameter tuning, the ideal parameters to tune the model are as follows: criterion = ”log_loss”, n_estimators = 50, random_state = 42.

4.2. Feature Engineering and Feature Extraction

In this research, we have used feature engineering concept to create new features with the help of existing features. So that, we can train the model with new set of features, rather than other models that are all using the existing features. So, we have created four new features of risk ratios, by calculating the risk percentage from the existing columns. Here are the newly created columns:

- age, we have changed the age to categorical by splitting the age values into 4 different bins, young age, middle age, and old age.
- cholesterol risk flag, where cholesterol >than 240
- High Blood Pressure risk flag, where resting blood pressure is >than 140.
- fitness index = Thal/Age.
- old_peak_slope = old peak * slope.
- age_exangia = age* exercise angia.

We then proceeded to perform feature extraction to extract only required features. Also to not overfit the model by training with more feature information. We have implemented Recursive Elimination Feature Extraction to extract the required features. After extracting the require features we then re-implemented the model to check it's performance. The model performance has slightly improved from 60.2% accuracy to 62% after feature extraction and 66% after hyperparameter tuning.

4.3. Model Implementation

In this dataset, we have mixture of categorical and numerical rows, we have also created new categorical features from the existing numerical features. So, we have aimed to use a non-parametric model and we have selected random forest classifier to classify our dataset. The dataset is slightly complex and have many hidden relationship between the features, and there are few outliers in the dataset, considering all these parameters we implemented random forest model. The base-line model got 60.2% accuracy. Further we have tuned the model by performing hyperparameter tuning, feature engineering and feature extraction and imiproved the model accuracy to 66%

4.4. Boosting

In this research we are using the random forest model as the base-line model and in order to improve the model performance, we implemented boosting technique by bringing all the weak learners together to train the model and reduce the bias of the model. The boosting techniques also helps us to reduce the variance of the model by focusing only on the important areas of the dataset, instead of focusing on the entire dataset and increasing the variance.So, as an extension of our baseline model we implemented Ada boosting, and Gradient boosting technique. The Gradient boosting technique has outperformed Ada Boosting by improving the model performance by 82% accuracy.

5. Explainable AI Techniques

In this research, we have implemented the XAI techniques like SHAP explaining the feature contribution to the final classification. There are many black box models applied on the heart disease dataset, but since we are dealing with the medical data, it is very important to explain the reasoning behind the classification. This will increase the trust and reliability in the medical industry experts.

The below graph shows the contribution of the newly created features towards the model classification. We can see that the chest_pain feature has the highest contribution towards the model classification.

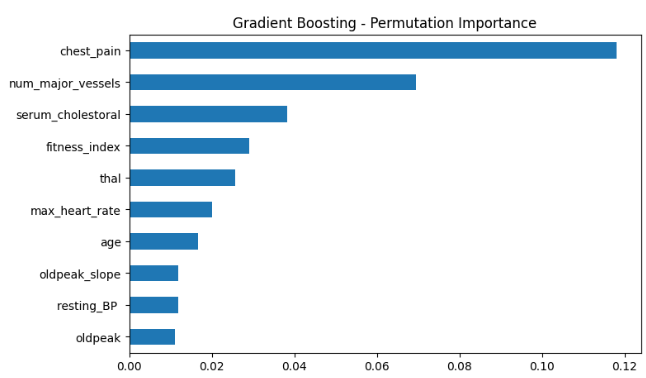

Figure 2. Explanation of Gradient boosting Model.

The below graph shows that for the ada boost model, the num_major_vessels feature has more contribution towards the final classification. Comparing with the gradient boosting model, it has equal importance along with the chest_pain feature. Along with the positive contributions we also have features that has negative contributions towards classification. Resting_blood_pressure has negative score for classifying a patient suffering from heart disease or not.

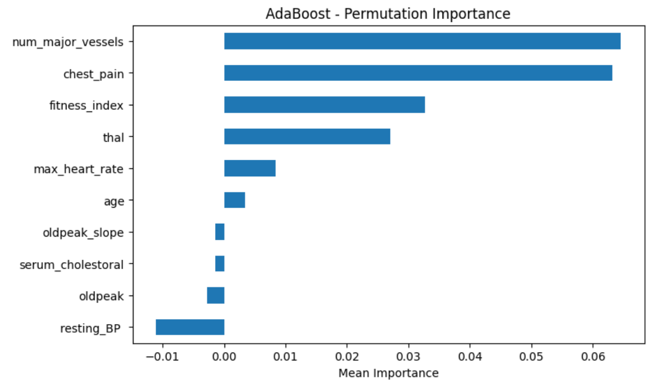

Figure 3. Explanation of Ada Boost Model.

For the Random forest model, there are very less features that has positive scores. This can mean that the random forest is not able to sub classify the patients into mild, severe, and very severe heart disease classes. For all the three models chest_pain is the only feature that has greater contribution for classifying the patients with heart disease or not.

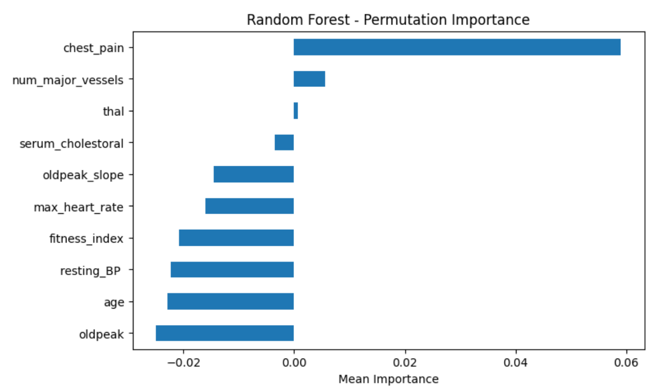

Figure 4. Explanation of Random Forest Classifier.

So, these Explainable AI techniques help the medical industry experts and researchers to focus on which features are very complex in terms of predicting the heart disease of the patient.

6. Evaluation

In this research, we have taken Precision, Recall, AUC and ROC curve metrics as our standard metrics to evaluate our model performance. This is a multi class dataset with 5 classes. Some of the classes were easy to classify but model was not able to find the hidden relationships in few classes.

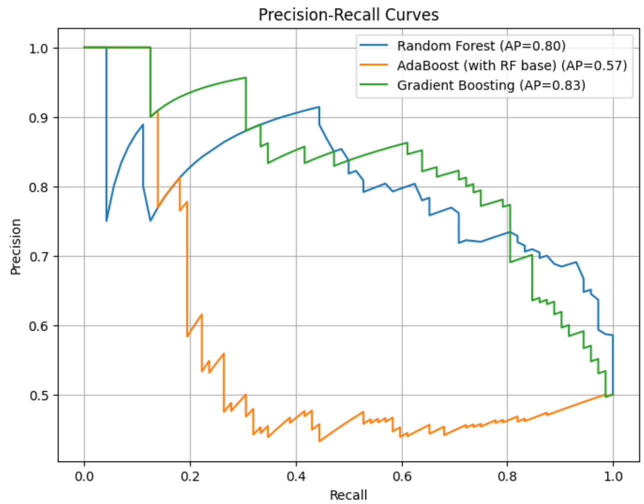

Figure 5. Precision Recall Curves

- Random Forest Classifier: The Baseline random forest classifier achieved 90% of recall score for the 0 class. That means it is able to predict very well with the patients that do not contain heart disease. Overall the F1 score is 76% which shows a balanced performance of the model but the downside of the model is that, it was not able to classify the 2 and 3 classes. The PR AUC score is 80% which is a very good performance.
- Ada Boost classifier: The Ada boost model has slightly lesser performance than random forest classifier but it has got reasonable results for the classes 0 and 1 with 75% and 69% accuracy. The Ada boost classifier has lower PR AUC score when compared to all the three models with the score of 0.565. This model is not able to classify the positives of the target variable.
- Gradient Boost classifier: The gradient boosting model has the highest performance in all the evaluation metrics. It has highest recall value for both 0 and 1 classes with the score of 88% and 64% respectively. It is able to predict the class 1 records better than random forest model. The weighted f1 score is 80% which is the highest in all the three models.

7. Conclusion and Future Work

In this research, we have taken a novel approach of creating new features that has not been explored in any other research and achieved a good F1 score of 80%. We have build computationally efficient models to classify the target variables. The future research can be focussed on building a model that can be able to classify the class 2 and 3 records as well. Currently we are able to distinguish the patients between no heart disease (0) and heart disease (1,2,3, and 4) but the further classification of mild, moderate, severe, and very severe heart diseases are not being classified. So, the future work of this study can be creating features that helps the model to understand the realationships between those classes and improving the performance of classification. Researchers can also fine tune the models to explore the hidden trends in the newly created features based on the SHAP outputs.

8. Acknowledgement

We would like to thank our supervisor, Dr. Zohaib Ijaz and Dr. Abhishek Kaushik, whose expertise was extremely valuable in formulating research questions and methodology. Your insightful feedback pushed us to sharpen my thinking and to bring our work to a greater position.

References

[1] V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *2020 2nd international conference on advances in computing, communication control and networking (ICACCCN)*, IEEE, 2020, pp. 177–181.

[2] N. Absar, E. K. Das, S. N. Shoma, *et al.*, "The efficacy of machine-learning-supported smart system for heart disease prediction," in *Healthcare*, MDPI, vol. 10, 2022, p. 1137.

[3] P. Guleria, P. Naga Srinivasu, S. Ahmed, N. Almusallam, and F. K. Alarfaj, "Xai framework for cardiovascular disease prediction using classification techniques," *Electronics*, vol. 11, no. 24, p. 4086, 2022.

[4] H. Mehta and K. Passi, "Social media hate speech detection using explainable artificial intelligence (xai)," *Algorithms*, vol. 15, no. 8, p. 291, 2022.

[5] L. Mhamdi, O. Dammak, F. Cottin, and I. B. Dhaou, "Artificial intelligence for cardiac diseases diagnosis and prediction using ecg images on embedded systems," *Biomedicines*, vol. 10, no. 8, p. 2013, 2022.

[6] R. Rajendran and A. Karthi, "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers," *Expert Systems with Applications*, vol. 207, p. 117 882, 2022.

[7] N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization," *Processes*, vol. 11, no. 4, p. 1210, 2023.

[8] A. M. Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective feature engineering technique for heart disease prediction with machine learning," *IEEE Access*, vol. 11, pp. 56 214–56 224, 2023.

[9] N. G. Rezk, S. Alshathri, A. Sayed, E. El-Din Hemdan, and H. El-Behery, "Xai-augmented voting ensemble models for heart disease prediction: A shap and lime-based approach," *Bioengineering*, vol. 11, no. 10, p. 1016, 2024.

[10] A. Sethi, S. Dharmavaram, and S. Somasundaram, "Explainable artificial intelligence (xai) approach to heart disease prediction," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, IEEE, 2024, pp. 1–6.

[11] D. Shrestha, "Advanced machine learning techniques for predicting heart disease: A comparative analysis using the cleveland heart disease dataset," *Applied Medical Informatics*, vol. 46, no. 3, 2024.

[12] B. Ahmad, J. Chen, and H. Chen, "Feature selection strategies for optimized heart disease diagnosis using ml and dl models," *arXiv preprint arXiv:2503.16577*, 2025.

[13] H. Lee and P. Tsoi, "Feature-enhanced machine learning for all-cause mortality prediction in healthcare data," *arXiv preprint arXiv:2503.21241*, 2025.