Journal of Information Systems Engineering and Management
2025, 10 (11s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Exploring Highly Influential Features and Model Comparison for Predicting Coronary Heart Disease: A Novel Approach Using SHAP

Dr. S. Meenakshi[1], Mrs. A. Chandra[2]
1 Assistant Professor, Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Potheri, Kattankukulathur, Tamil Nadu, India.-603203*
*1*Email: meenakshisankar2013@gmail.com.*
2 Assistant Professor, Department of Computer Application, Faculty of Science and Humanities, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College For Women. Vaishnava College Road, Shanthi Nagar, Chromepet, Chennai, Tamil Nadu, India.*

ARTICLE INFO

Received: 18 Nov 2024

Revised: 24 Dec 2024

Accepted: 12 Jan 2025

ABSTRACT

Forecasting coronary heart disease (CHD) is essential for improving healthcare outcomes and holds significant importance in identifying individuals at risk and mitigating severe health outcomes. Many prior studies have analyzed complete datasets without isolating the most influential features, which hinders the interpretability and efficacy of classification models. This research focus to address this gap by identifying key features that have been underexplored in past research on heart disease and evaluating multiple classification models based on these significant features. Using SHapley Additive exPlanations (SHAP), the most impactful features were extracted from a dataset comprising 1888 instances with 14 attributes and two outcome classes (Yes and No), indicating CHD risk. Key features such as 'thal,' 'chest pain type (cp),' The slope of the ST segment as shown during maximal exertion, '(slope)', 'ca,' and 'oldpeak' were prioritized for further analysis. Classification models, such as XGBoost, Decision Tree, K-Nearest Neighbours (KNN), and Logistic Regression were then assessed using this reduced feature set. Among these, XGBoost achieved the highest performance, with an 90.21 percent accuracy, 90.05% precision, 90.53% recall, and 90.29% F1 score. KNN, with an accuracy of 89.68%, came in second while the Decision Tree also yielded strong results with 90.21% accuracy. By leveraging a focused subset of critical features, this study demonstrates how classification models can be improved to produce results that are easier to understand and more effective. These findings pave the way for the development of a prototype tool to aid in heart disease detection, providing value to both research and clinical applications in cardiovascular diagnostics.

Keywords: Coronary Heart Disease (CHD), SHapley Additive exPlanations (SHAP), Classifier Model, Feature Importance, Feature Selection.

INTRODUCTION

One important subtype of cardiovascular disease (CVD), coronary heart disease (CHD), is a major global health concern and responsible for millions of deaths annually and contributing to a substantial proportion of worldwide mortality [1], [2]. Early prediction and diagnosis of CHD are vital for mitigating severe health complications and enhancing patient outcomes. Nevertheless, conventional diagnostic approaches, which depend on noticeable symptoms and static medical records, often fail to identify asymptomatic cases, thereby limiting their overall effectiveness [3][4].

Advances in machine learning (ML) have transformed healthcare by facilitating the development of predictive models capable of analysing complex interactions between features and outcomes [1][5]. These models utilize extensive datasets to improve prediction accuracy and provide interpretability, delivering insights that surpass traditional diagnostic approaches [6][7]. In the context of CHD prediction, feature selection methods like L1 regularisation, SHapley Additive exPlanations (SHAP) and SelectKBest have demonstrated their effectiveness in identifying critical features while minimizing computational demands [8][9].

Despite significant advancements, two key gaps remain in the existing research:

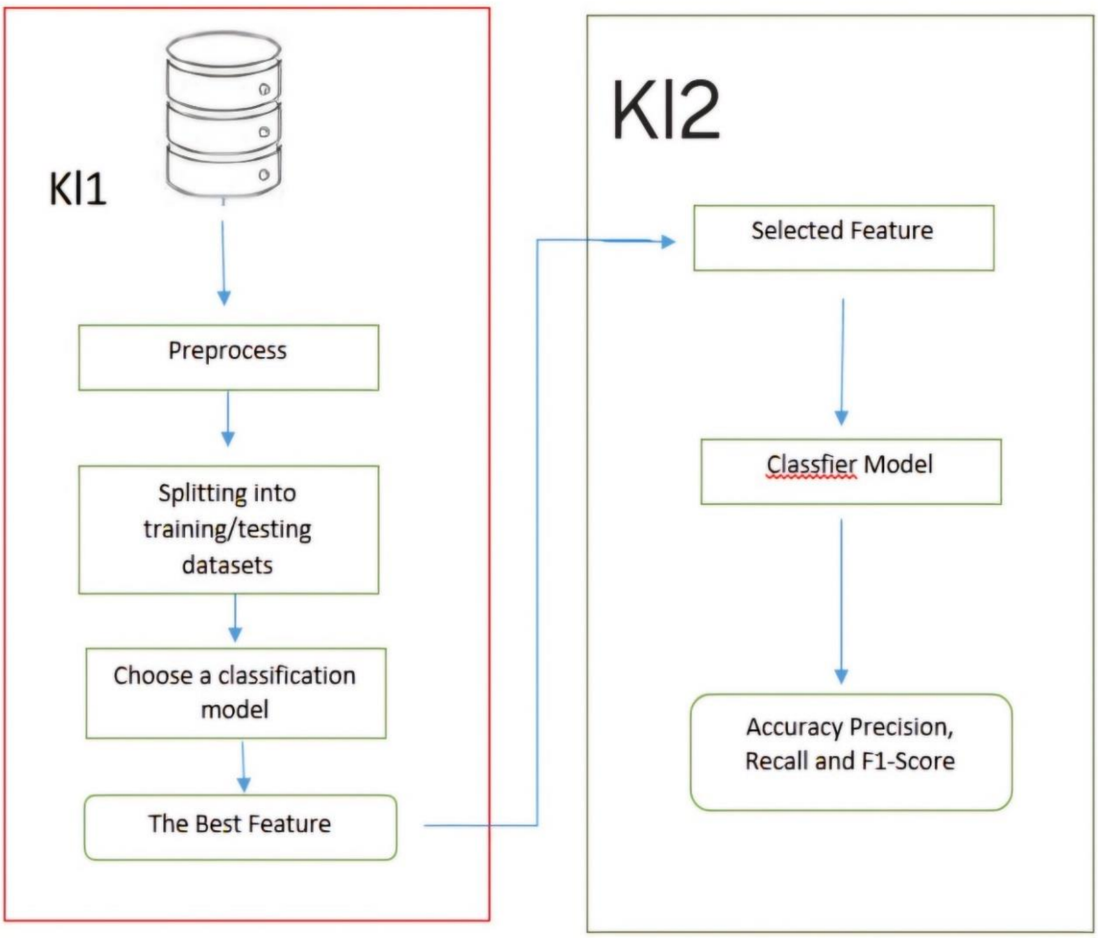
a. Key Identification 1 (KI1) – Insufficient Use of Feature Selection Models for CHD Prediction: Many studies focus on evaluating individual or ensemble models without employing feature selection techniques specifically tailored for CHD [6] [10].
b. Key Identification 2 (KI2) – Limited Integration of Selected Features in Classification Models: Most classification models do not adequately incorporate selected features, resulting in reduced interpretability and efficiency [11][12].

OBJECTIVES

To enhance the accuracy and interpretability of coronary heart disease (CHD) prediction models by identifying the most influential features using SHapley Additive exPlanations (SHAP) and evaluating the performance of classification models—XGBoost, Decision Tree, K-Nearest Neighbours (KNN), and Logistic Regression—based on the selected features.

METHODS

This research adopts a structured two-phase approach to improve the clarity and performance of classifier models for coronary heart disease (CHD) prediction as illustrated in Figure 1.



Phase 1: Feature Selection

The first phase focuses on Key Investigation 1 (KI1), which aims to determine the most significant features for predicting CHD. SHapley Additive exPlanations (SHAP) is employed to identify key attributes, including 'thal,' 'cp,' 'slope,' 'ca,' and 'oldpeak.' These features are essential for enhancing both the accuracy and interpretability of the models.

Phase 2: Model Training and Evaluation

In the second phase, the selected features are utilized to train various classification models such as XGBoost, K-Nearest Neighbours (KNN), Decision Trees, and Logistic Regression. Traditional criteria such as accuracy, precision, recall, and F1 score are used to evaluate the models to determine their effectiveness. This evaluation identifies the most suitable model for CHD prediction based on the selected features.

By integrating effective feature selection with comprehensive model evaluation, this two-phase approach addresses significant gaps in current research. It ensures a balance between interpretability and predictive performance, providing a more efficient and clinically meaningful framework for CHD detection.

Dataset

This phase of the analysis emphasizes identifying the categories of data required to meet the goals of the study, guaranteeing that the chosen data is both high-quality and aligned with the study's goals. The dataset outlined in Table 1 is directly relevant to the research topic is is available at https://github.com/trojrobert/Classification-of-heart-disease-uci-data-/blob/master/heart.csv. The GitHub Heart Disease Dataset is an open-source repository designed for public access, ensuring that the research data remains consistently updated and widely available. This dataset contains over 1,888 instances and 14 features, with its primary purpose being to predict the likelihood of a patient developing coronary heart disease**.**

Table 1: Heart Disease Dataset Attributes

Source: Adapted from GitHub Repository

S. No	Attribute Name	Feature Type	Description
1	Age	Integer	Age of the individual in years.
2	Sex	Binary	Gender of the individual: 1 for male, 0 for female.
3	Chest Pain Type	Integer	Categories of chest pain: 0 = typical angina, 1 = atypical angina, 2 = non-anginal pain, 3 = asymptomatic.
4	Resting BP	Integer	Resting blood pressure in millimeters of mercury (mmHg).
5	Cholesterol	Integer	Serum cholesterol level measured in milligrams per deciliter (mg/dL).
6	Fasting Blood Sugar	Binary	Indicates high fasting blood sugar levels: 1 = yes, 0 = no.
7	Resting ECG	Integer	Results from the resting electrocardiogram test.
8	Thalach	Integer	Maximum heart rate achieved during the test.
9	Exang	Binary	Exercise-induced angina: 1 = yes, 0 = no.
10	Oldpeak	Integer	ST segment depression caused by exercise relative to rest.
11	Slope	Integer	Slope of the ST segment during peak exercise: 1 = upsloping, 2 = flat, 3 = downsloping.
12	Ca	Integer	Number of major blood vessels (0–3) visible through fluoroscopy.
13	Thal	Integer	Thallium stress test results: 3 = normal, 6 = fixed defect.
14	Target	Binary	Indicates heart condition: 0 = no disease, 1 = disease.

Preprocessing:

A vital step in reducing mistakes and getting data ready for efficient analysis with machine learning models is data preprocessing. This process involves several interconnected steps tailored to the particulars of the dataset and the specifications of the chosen model. The primary goal is to ensure data accuracy and readiness for further analysis and model development.

The preprocessing begins with an inspection and cleaning phase, where the dataset is examined and refined by eliminating substantial missing values in rows or columns. This step is essential for ensuring the dataset is reliable, error-free, and suitable for analysis or modeling. Missing values can arise from data entry errors, typographical mistakes, or inconsistencies during data collection or downloading.

To detect missing values, the command data. Is null (). Sum () is utilized. This syntax provides a count of missing entries for each feature in the dataset. As shown in Figure. 2a, the results confirm that the dataset used in this study is complete, with no errors or inconsistencies identified during this stage.

```
Dataset loaded successfully!

Missing Values per Feature:
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalachh    0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

Figure 2a: Detecting Missing Values

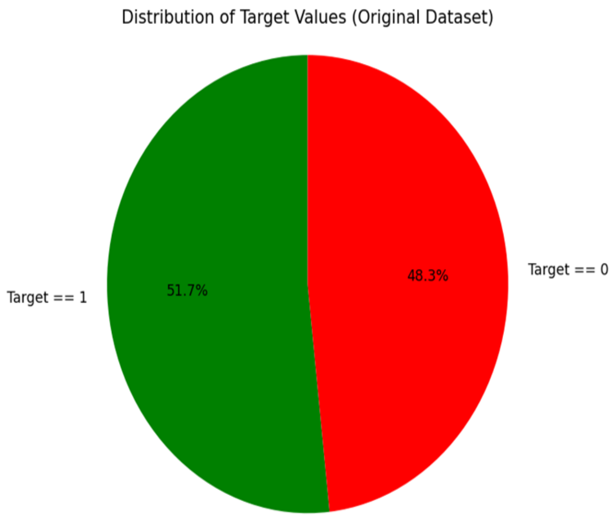


Figure 2b: Class Balancing:

The second stage of data preprocessing involves class balancing, which ensures an even distribution of instances for every dataset class or identifier. This step is particularly important for classification tasks, as Biassed models that support the majority class can result from unbalanced datasets, ultimately reducing predictive accuracy. For instance, if cases of heart disease are significantly fewer than non-heart disease cases, the model may struggle to correctly predict the minority class. Addressing such imbalances is critical for enhancing model robustness and accuracy.

In this study, the dataset comprises 1,888 records, with 911 samples labelled as target == 0 (48.3%) and 977 samples labelled as target == 1 (51.7%). This distribution indicates a reasonably balanced dataset, as the difference between the two classes is minimal. The accompanying figure 2b presents this distribution as a pie chart, where the red segment represents target == 0 ("No Heart Disease") and the green segment represents target == 1 ("Heart Disease"). The near-equal proportions of the two classes confirm that the dataset does not exhibit significant class imbalance, making additional balancing techniques unnecessary.

Splitting

Categorical features in a dataset represent unique groups or classes instead of numbers. These features are frequently expressed as labels or attributes that explain the type, condition, or classification of an item or instance. A common approach to handling such categorical data is one-hot encoding, which converts the categories into a binary representation. This technique creates separate columns for each category, assigning a value of 1 to indicate the presence of a particular category in a given observation, while all other columns are set to 0. For instance, as shown in Table 2, each categorical feature is transformed into binary columns, with a value of 1 indicating the corresponding category for a specific instance.

Table 2: Categorical Features

S.No	Attribute Name	Feature Type	Description
1	Sex	Binary	Denotes gender: 1 for male, 0 for female.
2	Fasting Blood Sugar	Binary	Indicates if the patient's fasting blood sugar level is high (1) or normal (0).
3	Exang	Binary	Angina brought on by exercise: 1 for presence, 0 for absence.
4	Target	Binary	Represents cardiac functionality: 0 for no disease, 1 for the presence of heart disease.

Numerical values are quantitative data attributes indicated by numbers that can be calculated, analyzed, or modified using mathematical operations. These features can be categorized as either discrete or continuous. Continuous features, in particular, are capable of taking any value within a specified range or interval. Examples of such features include variables like height, weight, and age, which can span an infinite spectrum and transition seamlessly among any two neighbouring values. The numerical features listed in Table 3 illustrate this concept.

Table 3: Numerical Features

S. No	Attribute Name	Feature Type	Description
1	Age	Number	Age of the individual in years.
2	Chest Pain Type	Number	Typical angina is represented by a score of 0, atypical angina by a score of 1, non-anginal pain by a score of 2, and asymptomatic by a score of 3.
3	Resting BP	Number	Resting blood pressure in mmHg
4	Cholesterol	Number	Serum Cholesterol levels measured in mg/dL for individuals without symptoms (asymptomatic).
5	Resting ECG	Number	Resting Electrocardiogram results
6	Thalach	Number	Attained maximum heart rate
7	Oldpeak	Number	Exercise-induced ST segment depression in comparison to rest
8	Slope	Number	Slope of the ST segment during peak exercise: Upsloping is represented by 1, flat by 2, and down sloping by 3.
9	Ca	Number	Count of principal blood vessels (numbers 0–3)visible through fluoroscopy.
10	Thal	Number	Results of the Thallium Stress Test: 3 indicates normal, 6 signifies a fixed defect.

Interpretable Model

To address Key Investigation 1 (KI1), which focuses on identifying the most critical factors in medical records for predicting coronary heart disease (CHD), SHapley Additive exPlanations (SHAP) was utilized for feature selection. SHAP, a game-theory-based framework, is well-known for measuring each feature's contribution to machine learning models, providing a global understanding of their impact on predictions.

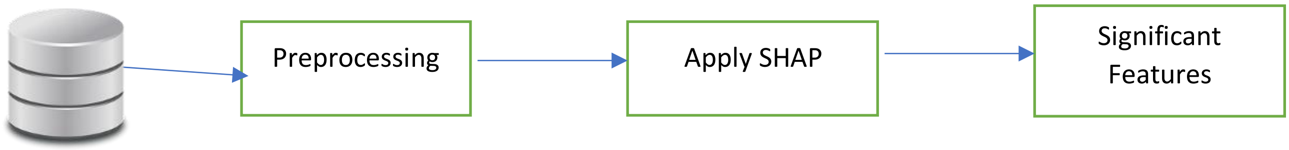


Figure 3: Interpretable Model

As illustrated in the Figure 3 above, the process begins with data preprocessing, followed by the application of SHAP to identify significant features. SHAP values are computed for each feature by analyzing their contribution to deviations from a baseline or average prediction value. The following formula is used to determine the SHAP value for a given feature i:

$$\phi_i = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

Where:

N represents the complete set of attributes.

A subset of features that do not include i is called S.

The model predicts f(Su{i}).

when feature i is included with subset *S*

This formula ensures an equitable allocation of the contributions from each feature by considering both individual effects and interactions among features.

Using SHAP, the study identified key characteristics like thal, type of chest pain (cp), and slope of segment during peak exercise (slope), ca, and oldpeak as the most influential factors for CHD prediction. These selected attributes were then employed to train and evaluate various classification models, enhancing their interpretability and predictive performance.

This feature selection technique not only boosts model accuracy but also maintains clinical relevance, offering healthcare professionals a clearer framework to assess CHD risk factors effectively.

Key Features

Identifying critical features is vital for improving the accuracy and interpretability of predictive models for coronary heart disease (CHD). In this study, the most impactful features were identified using SHapley Additive exPlanations (SHAP). These selected features were then utilized to evaluate various classification models, focusing on the most relevant attributes for CHD risk prediction.

Rather than analyzing all possible features, this process involved exploratory testing to prioritize those with the highest significance as determined by SHAP. By narrowing the feature set, the approach aimed to enhance both the predictive performance and the interpretability of the models.

The selected features—thal, chest pain type (cp), ST segment slope during maximal exertion (slope), ca, and oldpeak—were instrumental in optimizing the prediction process, contributing to more efficient and accurate CHD detection.

Classifier Model

Key Investigation 2 (KI2) seeks to identify the best classification model for predicting at predicting the likelihood of coronary heart disease (CHD) based on the most influential features. In this study, Decision trees (DT), K-Nearest Neighbours (K-NN), and logistic regression (LR), and XGBoost were evaluated. These models were chosen for their proven effectiveness in supervised learning tasks, particularly for heart disease prediction.

The dataset was separated into two subsets for evaluation: training data and testing data. The model was developed using the training data to find correlations between the input features (independent variables) and the target variable (dependent variable). The model's performance was then assessed using the testing data by assessing its predictive accuracy on previously unseen data.

This methodology ensures that the selected classification model is capable of generalizing effectively, making it suitable for real-world applications in predicting CHD risk.

Evaluation of Classifier Performance Using Key Metrics

Using important assessment criteria, such as accuracy, precision, recall, and F1-Score, the classifier model's performance was evaluated.to identify variations in performance. Once the test results were obtained, they were compared to identify the top-performing model for predicting heart disease, based solely on the most impactful features identified through SHAP.

RESULTS AND DISCUSSION

After the evaluation of the models, we examined the impact of each feature on the predictions, with a particular focus on the visual representation and interpretability of the variables. Figure 4a illustrates the influence of all the features on the model values using a color gradient, where blue represents a low contribution and red signifies a high contribution. The most influential features, including 'thal,' 'cp,' 'slope,' 'ca,' and 'oldpeak,' were identified as the top-ranking variables, as shown in Figure 4b. Additionally, Figure 4c highlights the mean SHAP values, providing a quantitative summary of the overall importance of each feature. The transition from blue to red in Figure 4a signifies an increasing influence of features on the predicted outcomes, while the ranking and mean values in Figures 4b and 4c guided the selection of the most significant features. By focusing on these key variables, the model aimed to increase accuracy, precision, and overall efficiency while minimizing time for processing and excluding less relevant features.

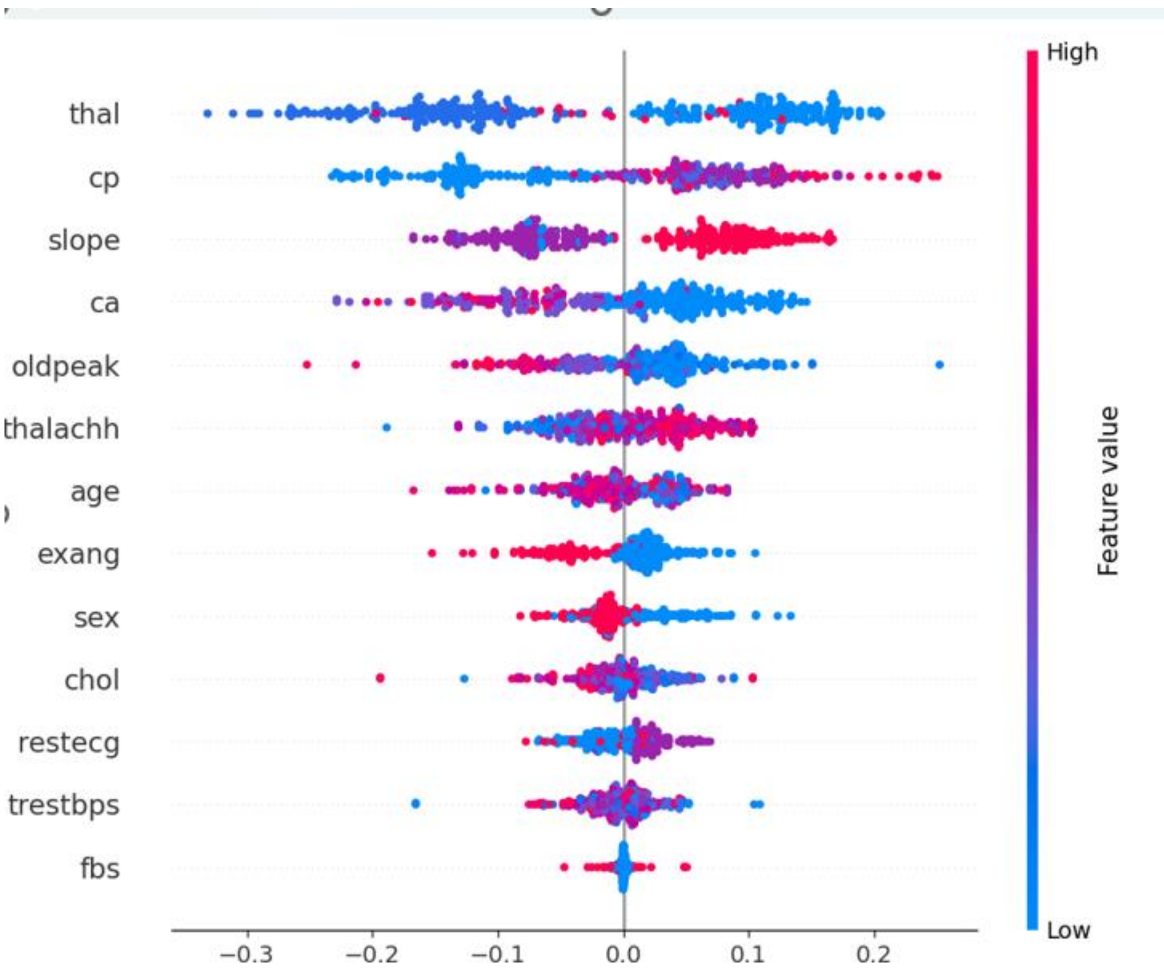


Figure 4a: SHAP value on all features

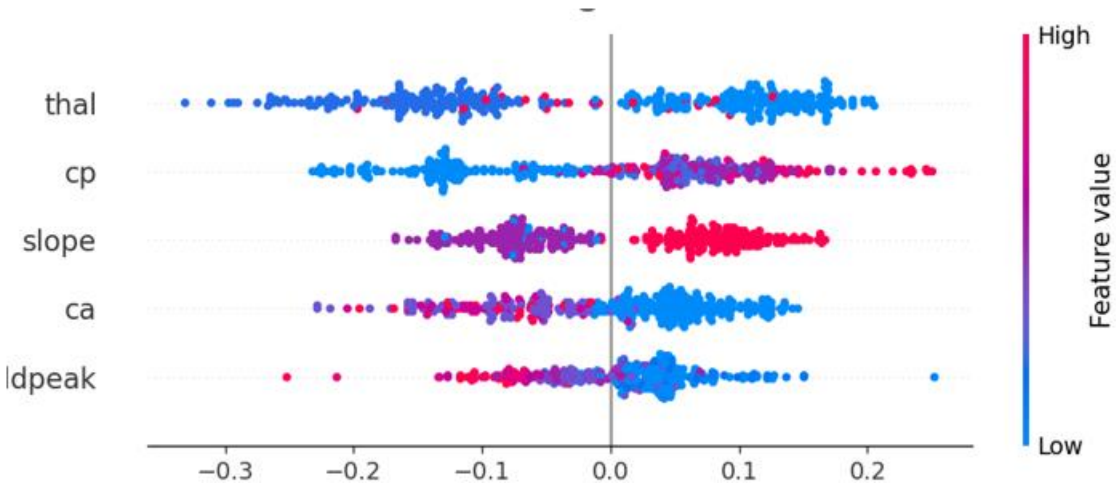


Figure 4b: Ranking of top 5 features

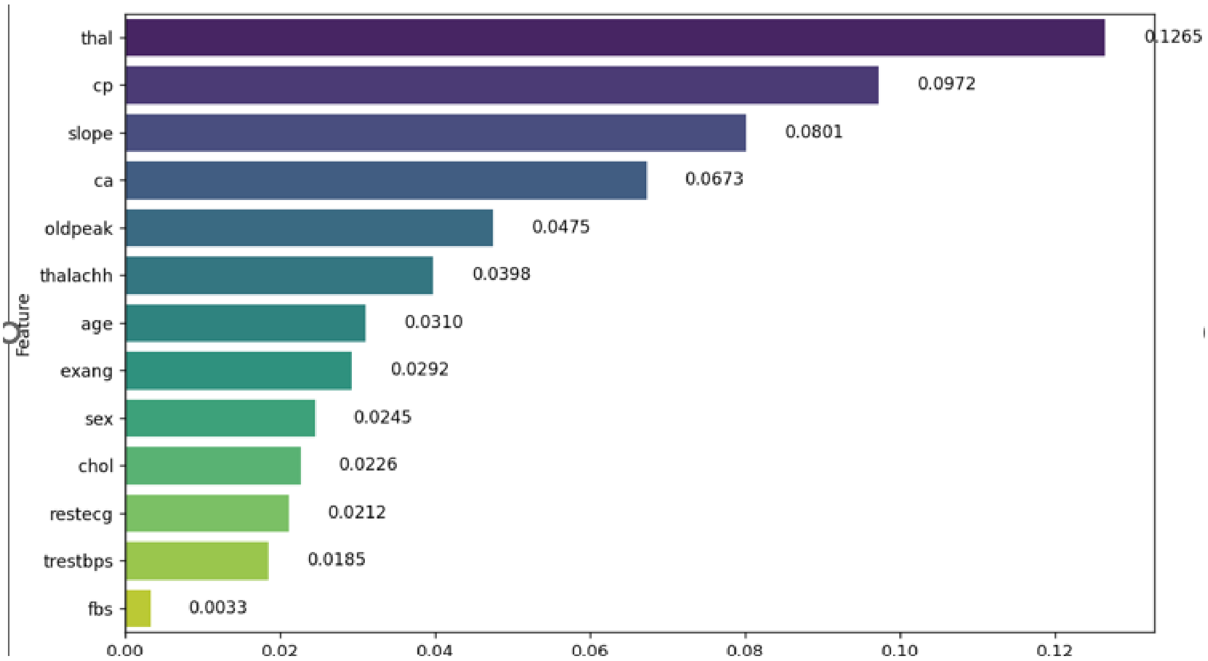


Figure 4c : MEAN SHAP values

As illustrated in Figure 5, which presents the performance metrics of the selected classifiers, Logistic Regression demonstrated the lowest performance across all metrics, achieving an accuracy of 0.6931. The notable disparity between precision and recall indicated the model's difficulty in distinguishing between classes, resulting in a relatively low F1 score of 0.7212. With an accuracy of 0.8968 and balanced precision (0.8832) and recall (0.9158), the K-Nearest Neighbours (KNN) model produced excellent results, resulting in an F1 score of 0.8992. With the highest accuracy of 0.9021, balanced precision (0.9005), and recall (0.9053), XGBoost was the best-performing model, resulting in an F1 score of 0.9029. The Decision Tree model performed comparably to XGBoost, also achieving an accuracy of 0.9021. Although it recorded the highest precision (0.9135), its slightly lower recall (0.8895) resulted in an F1 score of 0.9013, which was marginally below that of XGBoost. These results highlight how well tree-based models, especially XGBoost—perform, in predicting heart disease using the most significant features. Table 4 shows the summary of performance of selected classifierd

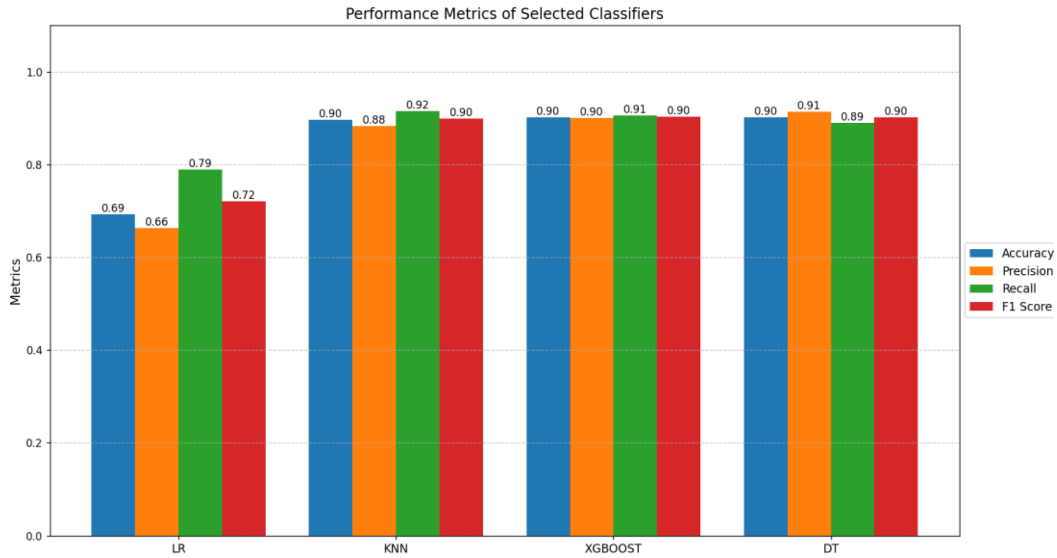


Figure 5: Performance Metrics of Selected Classifiers

Table 4: Classifier Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.6931	0.6637	0.7895	0.7212
KNN	0.8968	0.8832	0.9158	0.8992
XGBoost	0.9021	0.9005	0.9053	0.9029
Decision Tree	0.9021	0.9135	0.8895	0.9013

CONCLUSION

The SHAP analysis revealed that 'thal,' 'cp,' 'slope,' 'ca,' and 'oldpeak' were the most important characteristics for heart disease prediction. These features were then used to evaluate several classifier models, including Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and XGBoost. With an accuracy of 0.9021 and superior recall and F1 scores compared to the other models, XGBoost was the best performer.. The Decision Tree model also performed well, matching XGBoost in accuracy. These results suggest that tree-based models are particularly effective for this classification problem, likely due to their ability to capture intricate feature interactions and manage unpredictable relationships.

REFRENCES

[1] A. Supriyadi, A. Amali, A. T. Z. Y. Ahmad, D. K. Pramudito, A. Badruzzaman, and P. Purwanto, "The model interpretability on SHAP and comparison classification selection feature for heart disease prediction," Procedia Computer Science, vol. 245, pp. 210–219, 2024. doi: 10.1016/j.procs.2024.10.245.

[2] L. Ghoualmi, M. E. A. Benkechkache, and A. D. Ghoualmi, "Machine learning models for automated heart disease prognosis," in 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1–4. doi: 10.1109/IISEC56263.2022.9998233.

[3] B. Omarov, A. Batyrbekov, A. Suliman, B. Omarov, Y. Sabdenbekov, and S. Aknazarov, "Electronic stethoscope for detecting heart abnormalities in athletes," in 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1–5. doi: 10.1109/ACIT50332.2020.9300109.

[4] A. Mirsafaei and J. Basiri, "A novel data-driven method for cardiovascular disease prediction using ensemble learning," in 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 1235–1241. doi: 10.1109/BigData55660.2022.10001134.

[5] K. Shailaja, B. Seetharamulu, and M. A. Jabbar, "Machine learning in healthcare: A review," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 910–914. doi: 10.1109/ICECA.2018.8474918.

[6] M. Ayyanar, S. Jeganathan, S. Parthasarathy, V. Jayaraman, and A. R. Lakshminarayanan, "Predicting the cardiac diseases using SelectKBest method equipped Light Gradient Boosting Machine," in 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 117–122. doi: 10.1109/ICOEI53556.2022.9777224.

[7] S. Chua, V. Sia, and P. N. E. Nohuddin, "Comparing machine learning models for heart disease prediction," in 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2022, pp. 1–5. doi: 10.1109/IICAIET55139.2022.9936861.

[8] J. A. O'Reilly and W. Chanmittakul, "L1 regularization-based selection of EEG spectral power and ECG features for classification of cognitive state," in 2021 9th International Electrical Engineering Congress (iEECON), Pattaya, Thailand, 2021, pp. 365–368. doi: 10.1109/iEECON51072.2021.9440359.

[9] A. Dewan and M. Sharma, "Prediction of heart disease using a hybrid technique in data mining classification," in 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2015, pp. 704–706.

[10] F. Lopez-Caracheo, A. B. Camacho, C. A. Perez-Ramirez, M. Valtierra-Rodriguez, A. Dominguez-Gonzalez and J. P. Amezquita-Sanchez.(2018) "Fractal Dimension-based Methodology for Sudden Cardiac Death Prediction," IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, pp. 1-6.

[11] C. Boukhatem, H. Y. Youssef, and A. B. Nassif, "Heart disease prediction using machine learning," in 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, UAE, 2022, pp. 1–6. doi: 10.1109/ASET53988.2022.9734880.

[12] Ambesange, S., Vijayalaxmi, A., Sridevi, S. and Yashoda, B.S., 2020, July. "Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques." In *2020 fourth world conference on smart trends in systems, security and sustainability (WorldS4)* (pp. 827-832). IEEE. doi: 10.1109/WorldS450073.2020.9210404.

[13] M. Spann et al., "Applying machine learning in liver disease and transplantation: A comprehensive review," Hepatology, vol. 71, no. 3, pp. 1093–1105, 2020. doi: 10.1002/hep.31048.

[14] M. J. A. Jabbar and B. Samreen, "Heart disease prediction using hidden naive Bayes classifier," in 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada, 2016, pp. 396–400. doi: 10.1109/DSAA.2016.1234567.

[15] W. A. Bakar et al., "A review: Heart disease prediction in machine learning and deep learning," in 2023 19th IEEE International Colloquium on Signal Processing and Its Applications (CSPA), Kedah, Malaysia, 2023, pp. 150–155. doi: 10.1109/CSPA57446.2023.10087837.

[16] S. R. M. A. Ayeshmi M. and T. Peiris. (2022) "Heart Disease Stages Prediction using Machine Learning," 2022 8th International Conference on Big Data and Information Analytics (BigDIA), Guiyang, China, pp. 504-511

[17] K. L. Kumar and B. E. Reddy. (2021) "Heart Disease Detection System Using Gradient Boosting Technique," 2021 International Conference on Computing Sciences (ICCS), Phagwara, India, 2021, pp. 228-233.

[18] M. Bhatia and D. Motwani. (2020) "Use of Ensemblers Learning for Prediction of Heart Disease," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), Tirunelveli, India, pp. 1016-1023.

[19] P. Deepika and S. Sasikala, "Enhanced Model for Prediction and Classification of Cardiovascular Disease using Decision Tree with Particle Swarm Optimization," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1068-1072, doi: 10.1109/ICECA49313.2020.9297398.

[20] J. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer Informatics, vol. 2, pp. 59–77, 2006. doi: 10.1177/117693510600200030.

[21] J. A. Junaid and R. Kumar, "Data science and its application in heart disease prediction," in 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 396–400. doi: 10.1109/ICIEM48762.2020.9160056.

[22] K. S. L. Prasanna, N. P. Challa and J. Nagaraju. (2023) "Heart Disease Prediction using Reinforcement Learning Technique," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, pp. 1-5.

[23] Hassan, M. M., et al. (2023). "Performance evaluation of ensemble learning methods for coronary artery disease prediction." Journal of Medical Systems, 47(2), pp.1–12.

[24] Liu, H., et al. (2020). "Sparse feature selection for biomedical data classification." Bioinformatics, 36(3), pp.784–792.

[25] Chandrashekar, G., and Sahin, F. (2014). "A survey on feature selection methods." Computers & Electrical Engineering, 40(1), pp.16–28.