iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83848

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Explainable AI-Driven Heart Disease Prediction Using Ensemble Learning and SHAP-Based Clinical Interpretation

Sourav Angre[1], Ritesh Patil[2], Prasad Yeole[3], Mrs. Varsha Dharmadhikari[4]
Department of Information Technology, GH Raisoni College of Engineering and Management, Pune, India

Abstract: ***Heart disease remains one of the leading causes of mortality worldwide, emphasizing the need for accurate and interpretable predictive systems that can support early diagnosis and clinical decision-making. While machine learning techniques have demonstrated strong predictive capabilities in cardiovascular disease detection, their adoption in healthcare is often limited by the lack of transparency and explainability in model predictions. This study proposes an Explainable Artificial Intelligence (XAI)-driven framework for heart disease prediction using clin-ical parameters and ensemble learning techniques. The frame-work utilizes the UCI Heart Disease Dataset, comprising 1,025 patient records with 13 clinically relevant attributes, including age, chest pain type, cholesterol level, resting blood pressure, maximum heart rate achieved, and exercise-induced angina. Two machine learning models, namely Logistic Regression and Ran-dom Forest, are developed and evaluated for binary classification of heart disease risk. To enhance model transparency and clinical trust, SHapley Additive Explanations (SHAP) are integrated to provide both global and patient-level interpretations of model predictions. Global explanations identify the most influential clinical factors affecting prediction outcomes, while local explanations provide individualized reasoning for specific patient predictions. Ex-perimental results demonstrate that the Random Forest model outperforms Logistic Regression, achieving superior accuracy, precision, recall, and F1-score. Receiver Operating Character-istic (ROC) analysis further confirms the strong discriminative capability of the proposed framework, achieving an Area Under the Curve (AUC) of 0.857. SHAP-based analysis reveals that clinical attributes such as the number of major vessels (ca), chest pain type (cp), thalassemia status (thal), ST depression induced by exercise (oldpeak), and maximum heart rate achieved (thalach) are among the most influ-ential predictors of heart disease. The proposed framework not only delivers reliable predictive performance but also provides interpretable and clinically meaningful explanations, making it a practical decision-support tool for healthcare professionals.***

Index Terms: Explainable Artificial Intelligence, Heart Disease Prediction, Random Forest, Logistic Regression, SHAP, Clinical Decision Support System, UCI Heart Disease Dataset.

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain one of the leading causes of death worldwide and continue to pose a major challenge to public health systems. According to global health reports, millions of individuals are affected by heart-related disorders each year, resulting in significant mortality, health-care expenditure, and reduced quality of life. Early identifica-tion of patients at risk of developing heart disease is therefore essential for timely medical intervention, effective treatment planning, and improved patient outcomes. Traditional diag-nostic approaches rely heavily on clinical expertise, laboratory investigations, and medical imaging techniques. While these methods are effective, the increasing volume of healthcare data has created opportunities for data-driven approaches that can assist clinicians in making more accurate and timely decisions. Recent advancements in Machine Learning (ML) have enabled the development of predictive models capable of identifying complex relationships among clinical variables. Algorithms such as Logistic Regression, Support Vector Ma-chines, Decision Trees, and Random Forests have been widely employed for heart disease prediction using structured patient data. These models can analyze multiple clinical parameters simultaneously and identify patterns that may not be imme-diately apparent through conventional statistical analysis. As a result, machine learning-based systems have demonstrated promising performance in supporting cardiovascular risk as- sessment and disease prediction. Despite these advancements, the adoption of machine learn-ing in healthcare remains limited by concerns regarding in-terpretability and transparency. Many high-performing models function as black-box systems, generating predictions without providing clear explanations for their decisions. In clinical environments, healthcare professionals must understand the reasoning behind a prediction before incorporating it into patient care.

A lack of interpretability can reduce trust in au-tomated systems and create challenges in validating prediction outcomes. Consequently, there is a growing need for predictive frameworks that combine strong classification performance with meaningful explanations. Explainable Artificial Intelligence (XAI) has emerged as a promising solution to address these challenges. XAI tech-niques aim to improve transparency by providing insights into how machine learning models arrive at their predictions. Among the various XAI approaches, SHapley Additive Ex-planations (SHAP) has gained significant attention due to its strong theoretical foundation and ability to explain complex machine learning models. SHAP quantifies the contribution of individual features to a prediction, enabling both global interpretation of model behavior and local interpretation of individual patient outcomes. Such explanations are particu-larly valuable in healthcare, where understanding the factors influencing a diagnosis is often as important as the prediction itself.

This research proposes an Explainable AI-driven framework for heart disease prediction using clinical parameters obtained from the UCI Heart Disease Dataset. The framework integrates machine learning classification with SHAP-based explainabil-ity to provide both accurate predictions and interpretable insights. Logistic Regression and Random Forest models are developed and evaluated for binary classification, while SHAP is employed to identify the clinical factors contributing to prediction outcomes. By combining predictive analytics with explainability, the proposed system aims to support healthcare professionals in understanding model decisions and improving confidence in AI-assisted diagnosis.

The major contributions of this research are summarized as follows:

- Development of an explainable heart disease prediction framework using structured clinical data from the UCI Heart Disease Dataset.
- Comparative evaluation of Logistic Regression and Ran-dom Forest classifiers for cardiovascular disease predic-tion.
- Integration of SHAP-based explainability to provide both global feature importance analysis and patient-specific interpretations.
- Identification of clinically significant risk factors influ-encing heart disease prediction outcomes.
- Enhancement of transparency, trust, and interpretability in machine learning-assisted clinical decision-support sys-tems.

The remainder of this paper is organized as follows. Sec-tion II presents a review of existing literature related to heart disease prediction and explainable artificial intelligence. Section III describes the proposed methodology, including dataset preparation, model development, and SHAP-based explainability. Section IV discusses the experimental results and interpretability analysis. Finally, Section V concludes the paper and outlines potential directions for future research.

II. LITERATURE REVIEW

The application of computational techniques for cardio-vascular disease prediction has attracted significant attention over the past two decades. As heart disease continues to be a major cause of mortality worldwide, researchers have explored various statistical, machine learning, and artificial intelligence approaches to improve diagnostic accuracy and support clinical decision-making. The growing availability of electronic health records and structured clinical datasets has further accelerated the development of predictive models capa-ble of identifying patients at risk of developing cardiovascular conditions.

A. Traditional Heart Disease Prediction Approaches

Early heart disease prediction systems primarily relied on statistical analysis and rule-based diagnostic methods. Tech-niques such as logistic regression and multivariate statistical modeling were widely used to identify relationships between clinical risk factors and cardiovascular outcomes. These ap-proaches offered interpretability and ease of implementation; however, their predictive performance was often limited by assumptions of linear relationships among variables and their inability to effectively model complex interactions between clinical features. Medical practitioners traditionally relied on factors such as age, cholesterol levels, blood pressure, chest pain characteris-tics, and electrocardiographic findings to estimate cardiovascu-lar risk. Although these indicators remain clinically important, traditional statistical methods frequently struggle to capture the non-linear relationships that exist among multiple risk factors. Consequently, researchers began exploring machine learning techniques capable of extracting more sophisticated patterns from healthcare data.

B. Machine Learning-Based Heart Disease Prediction

Machine learning has emerged as a powerful tool for disease prediction due to its ability to analyze large datasets and identify hidden patterns within clinical information. Numerous studies have investigated the use of algorithms such as Logis-tic Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, Na¨ıve Bayes, and Artificial Neural Networks for heart disease classification.

Logistic Regression remains one of the most widely used baseline models because of its simplicity, computational effi-ciency, and interpretability. Several studies have demonstrated its effectiveness in predicting cardiovascular risk using struc-tured patient information. However, the model's linear nature often limits its ability to capture complex interactions among clinical variables.

Decision Tree-based approaches have gained popularity due to their intuitive structure and ability to model non-linear re-lationships. Similarly, Support Vector Machines have demon-strated strong classification performance in various healthcare applications, particularly when dealing with high-dimensional data. Neural network-based models have also been employed for cardiovascular disease prediction, leveraging their ability to learn complex feature representations. Despite achieving promising predictive accuracy, these methods often require extensive parameter tuning and may suffer from reduced interpretability.

Recent research has shown that machine learning techniques generally outperform traditional statistical approaches in dis-ease prediction tasks. Nevertheless, predictive performance alone is insufficient in healthcare environments, where clin-icians require explanations and justification for algorithmic decisions before integrating them into medical practice.

C. Ensemble Learning for Cardiovascular Risk Prediction

Ensemble learning techniques combine the predictions of multiple models to improve classification performance and generalization capability. Among these approaches, Random Forest has emerged as one of the most successful algorithms for healthcare prediction tasks. Introduced by Breiman, Ran-dom Forest constructs multiple decision trees using randomly selected subsets of data and features, reducing overfitting while improving robustness.

Several studies have reported that Random Forest consis-tently achieves high predictive performance in heart disease classification. The algorithm effectively handles heterogeneous clinical data, accommodates complex feature interactions, and remains relatively resistant to noise and missing values. These characteristics make it particularly suitable for medical datasets containing diverse patient attributes and varying risk profiles.

Comparative studies have frequently demonstrated that Ran-dom Forest outperforms traditional classifiers such as Logistic Regression and Decision Trees in terms of accuracy, precision, recall, and overall predictive capability. As a result, Random Forest has become a popular choice for cardiovascular risk assessment and disease prediction systems.

D. Explainable Artificial Intelligence in Healthcare

Although machine learning models can achieve high pre-dictive accuracy, their practical adoption in healthcare is often hindered by limited interpretability. Many advanced algorithms operate as black-box systems, providing predictions without explaining the reasoning behind them. This lack of trans-parency can reduce clinician trust and create challenges in validating model decisions.

To address these concerns, Explainable Artificial Intelli-gence (XAI) has emerged as an active area of research. XAI aims to improve transparency by providing interpretable insights into machine learning predictions. Among the vari-ous explainability techniques, SHapley Additive Explanations (SHAP) has gained widespread recognition due to its strong theoretical foundation in cooperative game theory and its ability to explain complex models consistently.

SHAP assigns contribution values to individual features based on their influence on prediction outcomes. The method supports both global explanations, which reveal the overall importance of features across a dataset, and local explanations, which provide patient-specific interpretations for individual predictions. Recent healthcare studies have demonstrated the effectiveness of SHAP in improving transparency, facilitating model validation, and enhancing clinician confidence in AI-assisted diagnostic systems.

E. Research Gap and Motivation

Despite significant advancements in machine learning-based heart disease prediction, several challenges remain. Many existing studies primarily focus on maximizing predictive ac-curacy while providing limited attention to model interpretabil-ity. Furthermore, some explainable systems offer only global feature importance analysis without generating individualized explanations that can support patient-level clinical decision-making.

The need for transparent and trustworthy predictive systems has become increasingly important as artificial intelligence moves toward real-world healthcare deployment. To address these limitations, this research proposes an Explainable AI-driven framework that combines the predictive strength of ensemble learning with SHAP-based interpretability. By in-tegrating Random Forest classification with both global and patient-level explanations, the proposed framework aims to deliver accurate, transparent, and clinically meaningful heart disease predictions that can support healthcare professionals in decision-making processes.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

III. PROPOSED METHODOLOGY

A. Dataset Description

The effectiveness of any machine learning-based healthcare prediction system largely depends on the quality and reliability of the dataset used for model development and evaluation. In this study, the UCI Heart Disease Dataset was utilized as the primary data source for heart disease prediction. The dataset is publicly available through the University of California, Irvine (UCI) Machine Learning Repository and is widely recognized as one of the most frequently used benchmark datasets for cardiovascular disease research. Due to its extensive use in predictive healthcare studies, the dataset provides a reliable foundation for evaluating machine learning algorithms and comparing results with existing literature.

The dataset contains a total of 1,025 patient records, each representing an individual patient evaluation. Every record consists of 13 clinical attributes and one target variable indi-cating the presence or absence of heart disease. The included features represent demographic characteristics, physiological measurements, laboratory findings, and cardiovascular risk indicators commonly used during clinical assessment. These attributes capture multiple aspects of patient health and provide valuable information for disease prediction.

Table I summarizes the clinical features used in this study. The target variable is represented as a binary classification label, where a value of 1 indicates the presence of heart disease and a value of 0 indicates the absence of heart disease. This

TABLE I
CLINICAL FEATURES USED FOR HEART DISEASE PREDICTION

Feature	Meaning
Age	Age
Sex	Gender
CP	Chest Pain Type
Trestbps	Resting BP
Chol	Cholesterol
FBS	Fasting Blood Sugar
Restecg	ECG Result
Thalach	Max Heart Rate
Exang	Exercise Angina
Oldpeak	ST Depression
Slope	ST Slope
CA	Major Vessels
Thal	Thalassemia

formulation allows the prediction problem to be treated as a supervised binary classification task. The objective of the proposed framework is to learn the relationship between the clinical attributes and the target label in order to accurately classify patients according to their cardiovascular risk.

The UCI Heart Disease Dataset was selected for this re-search due to several advantages. First, it contains clinically relevant features that are routinely collected during cardio-vascular examinations, making the findings more applicable to real-world healthcare settings. Second, the dataset has been extensively validated and used in previous heart dis-ease prediction studies, enabling meaningful comparison with existing machine learning approaches. Finally, its balanced representation of patient records and well-structured feature set make it suitable for evaluating both predictive performance and explainability techniques.

By utilizing a standardized and widely accepted dataset, the proposed framework ensures reproducibility, reliability, and comparability of results while providing a strong foundation for the development of explainable heart disease prediction models.

B. Data Preprocessing

Data preprocessing is a crucial stage in machine learning applications, as the quality and consistency of input data di-rectly influence model performance and prediction reliability. Healthcare datasets often contain missing values, inconsis-tencies, redundant information, or imbalanced distributions that can negatively affect the learning process. Therefore, an appropriate preprocessing strategy is essential to ensure that the data are suitable for model development and evaluation.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

In this study, the UCI Heart Disease Dataset was first inspected to verify data integrity and consistency. The dataset was examined for missing values, duplicate records, and invalid entries that could introduce bias into the predictive models. Since the selected dataset was well-structured and contained complete clinical information, extensive data clean-ing procedures were not required. Nevertheless, a systematic inspection process was performed to ensure the reliability of the experimental results.

Following data validation, the dataset was separated into input features and target labels. The thirteen clinical attributes, including demographic information, physiological measure-ments, and cardiovascular risk indicators, were designated as predictor variables. The target attribute representing the presence or absence of heart disease was extracted as the clas-sification label. This separation enabled the machine learning algorithms to learn the relationship between patient character-istics and disease outcomes.

To facilitate model training and evaluation, the dataset was partitioned into training and testing subsets using an 80:20 split ratio. Approximately 80 percent of the records were allocated for training, while the remaining 20 percent were reserved for independent testing. The training subset was utilized to learn predictive patterns from historical patient data, whereas the testing subset provided an unbiased assessment of model performance on previously unseen records. A fixed random state was employed during dataset partitioning to ensure reproducibility and consistency across experiments. The preprocessing workflow consists of dataset validation, feature-target separation, train-test partitioning, model train-ing, and performance evaluation.

The resulting preprocessed dataset provides a structured and reliable foundation for machine learning-based heart disease prediction. By ensuring data consistency and establishing a reproducible training and evaluation process, the preprocessing stage contributes significantly to the robustness and credibility of the proposed framework.

C. Model Development

The core objective of the proposed framework is to accu-rately classify patients according to the presence or absence of heart disease using clinically relevant attributes. To achieve this objective, two supervised machine learning algorithms, Logistic Regression and Random Forest, were employed and evaluated. These models were selected due to their widespread use in healthcare analytics, strong predictive capabilities, and suitability for binary classification problems.

1) *Logistic Regression:* Logistic Regression is one of the most commonly used statistical learning algorithms for binary classification tasks. The model estimates the probability of an event occurring by applying a logistic sigmoid function to a linear combination of input features. Due to its simplicity and interpretability, Logistic Regression is frequently used as a baseline classifier in healthcare prediction studies.

   For an input feature vector X, the probability of heart disease occurrence is computed using the logistic function:

$$P(Y=1|X)=\frac{1}{1+e^{-z}} \quad (1)$$

where

$$z=\beta_0+\sum_{i=1}^{n}\beta_i x_i \quad (2)$$

Although Logistic Regression provides interpretable results and computational efficiency, its ability to capture complex non-linear relationships among clinical variables is limited. Therefore, an additional ensemble learning approach was investigated.

2) *Random Forest:* Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve predictive performance and reduce overfitting. Each tree is constructed using randomly selected subsets of training data and feature attributes. The final prediction is obtained through majority voting across all decision trees within the ensemble.

The Random Forest prediction process can be represented as:

$$\hat{Y} = \text{mode}(T_1(X), T_2(X), \ldots, T_N(X)) \quad (3)$$

where $T_i(X)$ represents the output of the i-th decision tree and N denotes the total number of decision trees in the affect patient outcomes, understanding why a prediction is generated is often as important as the prediction itself. To address this challenge, the proposed framework incorporates SHapley Additive Explanations (SHAP), a widely adopted Ex-plainable Artificial Intelligence (XAI) technique that improves model interpretability. SHAP is based on the concept of Shapley values derived from cooperative game theory. The technique interprets ma-chine learning predictions by assigning contribution scores to individual features according to their influence on the final prediction. Each feature is treated as a participant in a coop-erative game, and its contribution is calculated by evaluating its impact across multiple feature combinations. This approach provides a theoretically consistent and mathematically sound explanation of model behavior.

The SHAP explanation model can be represented as:

$$f(x) = \phi_0 + \sum_{i=1}^{n} \phi_i \quad (4)$$

Random Forest model.

Random Forest offers several advantages for healthcare applications. The algorithm effectively handles heterogeneous clinical data, models non-linear feature interactions, and ex-hibits strong resistance to overfitting. Furthermore, its tree-based structure integrates naturally with explainability tech-niques such as SHAP, enabling detailed interpretation of prediction outcomes.

3) *Model Training and Prediction:* After preprocessing, the training subset was supplied to both classifiers to learn the relationship between clinical parameters and heart disease outcomes. The models were trained using the thirteen clinical attributes as predictor variables and the binary disease label as the target variable. Following training, predictions were generated for the testing dataset and evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and ROC-AUC.

Comparative evaluation was performed to determine the most suitable classifier for explainable heart disease prediction. Experimental results demonstrated that the Random Forest model consistently achieved superior predictive performance compared to Logistic Regression. Consequently, Random For-est was selected as the primary prediction model for subse-quent SHAP-based explainability analysis and clinical inter-pretation.

By combining a baseline statistical classifier with an en-semble learning approach, the proposed framework provides a comprehensive evaluation of predictive performance while en-suring that the selected model remains suitable for transparent and explainable healthcare decision-support applications.

D. SHAP Explainability

While machine learning models have demonstrated signif-icant success in healthcare prediction tasks, their adoption in real-world clinical settings is often limited by a lack of trans-parency. Many high-performing models, particularly ensemble and deep learning methods, operate as black-box systems that provide predictions without explaining the reasoning behind their decisions. In healthcare, where clinical decisions directly where $f(x)$ denotes the prediction produced by the model, ϕ_0 represents the baseline output, and ϕ_i quantifies the con-tribution of the i-th feature to the final prediction. Positive SHAP values increase the likelihood of heart disease, whereas negative SHAP values decrease the predicted risk.

One of the primary advantages of SHAP is its ability to pro-vide both global and local interpretability. Global explanations help identify the most influential features across the entire dataset, enabling researchers and healthcare professionals to understand the overall behavior of the prediction model. By analyzing feature importance across all patient records, it becomes possible to determine which clinical attributes have the greatest impact on heart disease prediction. Such insights support model validation and ensure that the classifier relies on medically meaningful information.

In addition to global interpretation, SHAP provides local explanations for individual predictions. Patient-level expla-nations reveal how specific clinical features contribute to a particular prediction outcome. This capability is especially valuable in healthcare environments, where clinicians often require justification for predictions generated for individual patients. Through local explanations, healthcare professionals can identify the factors responsible for elevated cardiovascular risk and evaluate whether the prediction aligns with established clinical knowledge.

SHAP was selected for this research due to several ad-vantages over alternative explainability methods. First, it pro-vides consistent feature attribution and maintains a strong theoretical foundation based on cooperative game theory. Sec-ond, it supports both global and patient-specific interpretation within a unified framework. Third, SHAP integrates effectively with tree-based machine learning algorithms such as Random Forest, allowing efficient generation of explanations without compromising predictive performance.

By incorporating SHAP into the proposed framework, the system extends beyond traditional classification and becomes an interpretable clinical decision-support tool. The generated explanations enhance transparency, improve trust in machine learning predictions, and enable healthcare professionals to un-derstand the clinical factors influencing prediction outcomes. Consequently, SHAP plays a central role in bridging the gap between predictive accuracy and practical clinical usability within the proposed heart disease prediction framework.

E. System Pipeline

The proposed framework integrates machine learning-based prediction and explainable artificial intelligence into a unified clinical decision-support workflow. The system is designed to process structured patient information, generate heart disease predictions, and provide interpretable explanations that assist healthcare professionals in understanding the reasoning behind model decisions.

Initially, patient clinical parameters are collected and sup-plied as input to the framework. These parameters include demographic information, physiological measurements, and cardiovascular risk indicators obtained from the UCI Heart Disease Dataset. The collected data undergo preprocessing and validation to ensure consistency and suitability for machine learning analysis.

Following preprocessing, the prepared feature set is pro-vided to the prediction module, where Logistic Regression and Random Forest classifiers are trained and evaluated. Based on comparative performance analysis, Random Forest is selected as the primary prediction model. The trained classifier analyzes patient attributes and generates a binary prediction indicating the presence or absence of heart disease.

To enhance transparency, the generated prediction is subse-quently processed by the SHAP explainability module. SHAP computes feature contribution scores and identifies the clinical parameters that most strongly influence prediction outcomes. The framework supports both global explanations, which re-veal overall model behavior, and patient-specific explanations, which provide individualized reasoning for a particular pre-diction.

The final stage of the pipeline focuses on clinical inter-pretation and decision support. The generated explanations enable healthcare professionals to identify key cardiovascular risk factors, validate prediction outcomes, and gain greater confidence in AI-assisted diagnosis. By integrating prediction and explainability within a single workflow, the framework provides both predictive accuracy and interpretability.

The overall workflow of the proposed system is illustrated in Fig. 1.

The integration of machine learning and explainable artifi-cial intelligence enables the proposed framework to function as an interpretable and reliable heart disease prediction system suitable for healthcare decision-support applications.

IV. RESULTS AND ANALYSIS

A. Performance Evaluation

The performance of the proposed Explainable AI-driven heart disease prediction framework was evaluated using the UCI Heart Disease Dataset. Following preprocessing, the dataset was divided into training and testing subsets using an 80:20 ratio. Two machine learning algorithms, Logistic Regression and Random Forest, were trained and evaluated to determine the most suitable classifier for heart disease pre-diction. Logistic Regression was selected as a baseline model due to its simplicity and interpretability, while Random Forest was employed as an ensemble learning approach capable of capturing complex relationships among clinical variables.

To comprehensively assess predictive performance, multiple evaluation metrics were utilized, including accuracy, precision, recall, F1-score, confusion matrix analysis, and Receiver Op-erating Characteristic (ROC) analysis. These metrics provide complementary perspectives on model effectiveness and en-able a balanced evaluation of classification performance. In healthcare applications, particular emphasis is placed on recall and F1-score, as these metrics reflect the model's ability to correctly identify patients with heart disease while minimizing misclassification.

Table II presents the comparative performance of Logistic Regression and Random Forest across the selected evaluation metrics.

TABLE II
PERFORMANCE COMPARISON OF CLASSIFICATION MODELS

Metric	Logistic Regression	Random Forest
Accuracy	0.873	0.947
Precision	0.860	0.950
Recall	0.840	0.930
F1-Score	0.850	0.940

The results indicate that the Random Forest classifier consis-tently outperforms Logistic Regression across all performance metrics. The Random Forest model achieved an accuracy of 94.7 percent, significantly exceeding the performance of Logistic Regression.

Similarly, improvements were observed in precision, recall, and F1-score, demonstrating the effectiveness of ensemble learning in modeling complex clinical relationships. The higher recall achieved by Random Forest is particularly important in healthcare settings because it reduces the likelihood of failing to identify patients who are genuinely affected by heart disease.

The comparative performance of the two classifiers is illustrated in Fig. 2.

As shown in Fig. 2, the Random Forest classifier consistently achieves higher scores across all evaluation metrics. The results suggest that the ensemble-based learning mechanism enables the model to better capture non-linear interactions among cardiovascular risk factors, resulting in improved predictive capability and robustness.

To further analyze classification performance, a confusion matrix was generated for the Random Forest classifier. The confusion matrix provides a detailed representation of pre-diction outcomes by illustrating the number of correctly and incorrectly classified instances.

The confusion matrix shown in Fig. 3 demonstrates that the majority of patient records were correctly classified into their

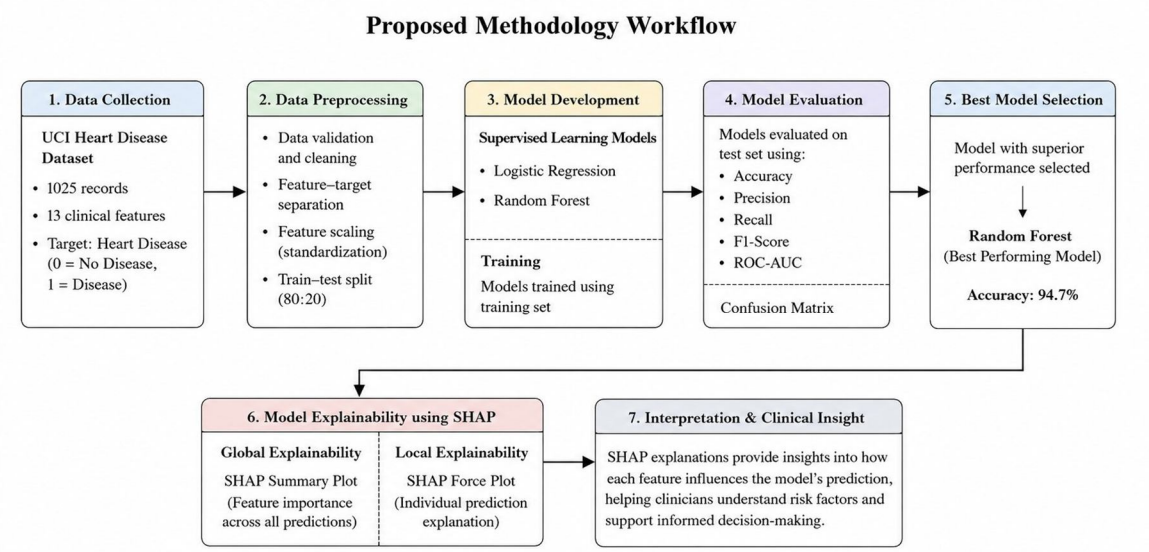


Fig. 1. Workflow of the proposed heart disease prediction framework using machine learning and SHAP explainability.

Fig. 1. Workflow of the proposed Explainable AI-driven heart disease prediction framework. Clinical parameters are preprocessed and supplied to machine learning models for prediction, followed by SHAP-based explanation and clinical interpretation.

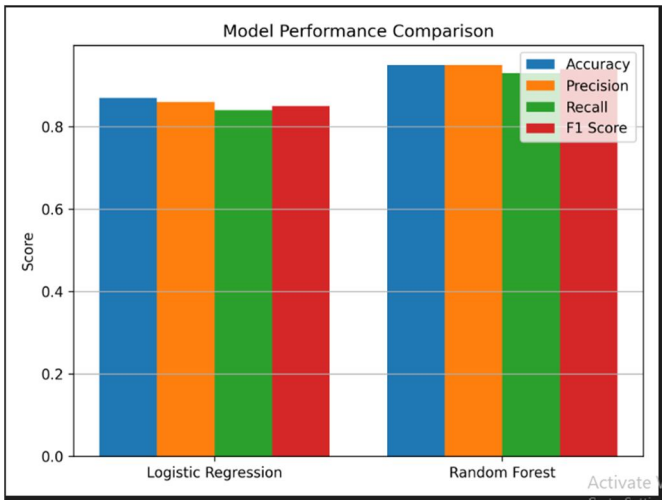


Fig. 2. Performance comparison between Logistic Regression and Random Forest using accuracy, precision, recall, and F1-score metrics.

respective categories. A high number of true positive and true negative predictions, combined with a relatively low number of false classifications, indicates that the Random Forest model possesses strong discriminative capability. Such performance is essential in clinical applications, where diagnostic errors can significantly affect patient outcomes.

In addition to classification metrics, Receiver Operating Characteristic (ROC) analysis was conducted to evaluate the model's ability to distinguish between positive and negative classes across varying decision thresholds. The ROC curve provides a graphical representation of the trade-off between

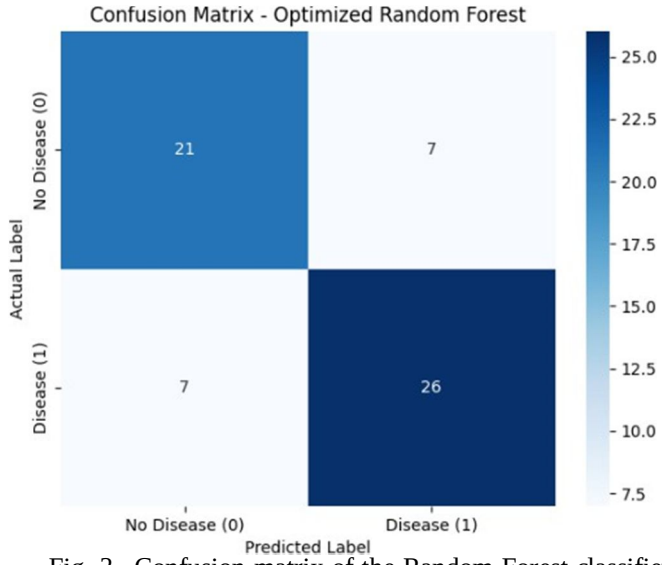


Fig. 3. Confusion matrix of the Random Forest classifier.

the True Positive Rate and False Positive Rate, while the Area Under the Curve (AUC) serves as a threshold-independent measure of predictive performance.

The obtained AUC value of 0.857 indicates strong classifi-cation capability and confirms that the model can effectively distinguish between patients with and without heart disease. An AUC significantly greater than 0.5 demonstrates that the classifier performs substantially better than random prediction and possesses meaningful predictive value for healthcare ap-plications.

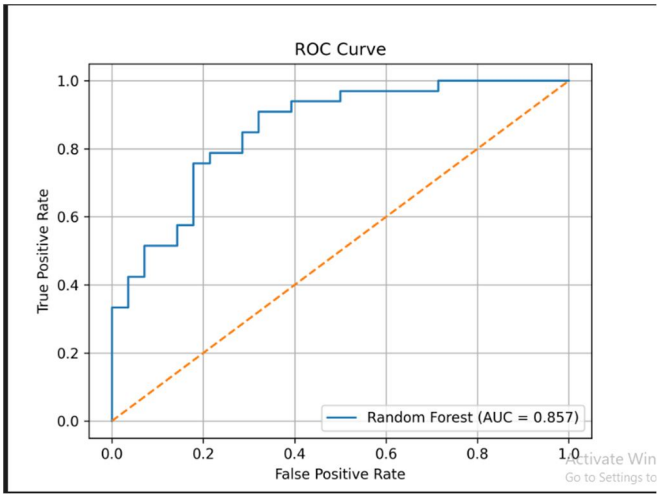


Fig. 4. ROC curve of the Random Forest classifier with an AUC value of 0.857.

Overall, the quantitative evaluation demonstrates that Ran-dom Forest provides superior predictive performance com-pared to Logistic Regression and serves as a reliable founda-tion for the explainability analysis presented in the subsequent sections. The combination of high accuracy, strong discrimina-tive capability, and robust classification performance supports the suitability of the proposed framework for heart disease prediction and clinical decision-support applications.

B. SHAP-Based Explainability

While the quantitative evaluation confirms the effectiveness of the Random Forest classifier in predicting heart disease, predictive accuracy alone is insufficient for healthcare appli-cations. Clinicians require explanations that justify prediction outcomes and identify the clinical factors influencing model decisions. To address this requirement, the proposed frame-work integrates SHapley Additive Explanations (SHAP) to provide transparent and interpretable insights into the predic-tion process.

SHAP enables the interpretation of complex machine learn-ing models by assigning contribution scores to individual features according to their influence on prediction outcomes. Unlike traditional feature importance methods that provide only aggregate importance values, SHAP offers both global explanations that describe overall model behavior and local explanations that justify individual patient predictions. This dual interpretability capability makes SHAP particularly suit-able for healthcare applications where both population-level understanding and patient-specific reasoning are essential.

1) *Global Feature Importance Analysis:* Global interpretation was performed using a SHAP summary plot, which ranks clinical features according to their overall contribution to heart disease prediction. The summary plot also illustrates the direction and magnitude of feature influence across all patient records, enabling a comprehensive understanding of model behavior.

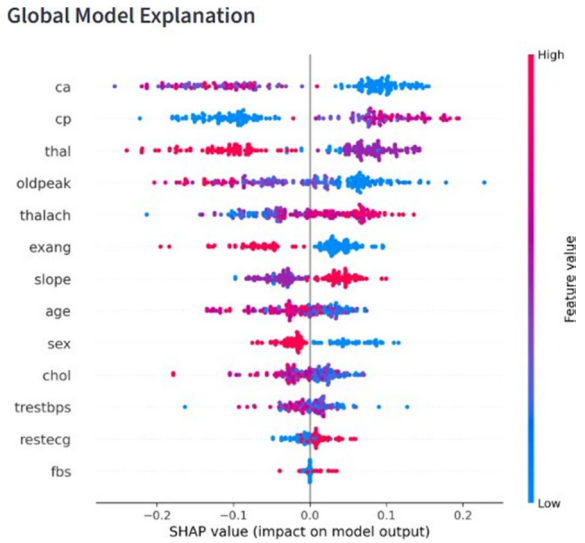


Fig. 5. SHAP summary plot showing global feature importance and the influence of clinical parameters on heart disease prediction.

As shown in Fig. 5, the number of major vessels visualized by fluoroscopy (*ca*) emerged as the most influential predictor of heart disease. This observation is clinically meaningful because increased vessel abnormalities are strongly associated with coronary artery disease and cardiovascular complications. Chest pain type (*cp*) also demonstrated substantial predictive influence, reflecting its importance as a primary diagnostic indicator in cardiovascular assessment.

Additional influential features include thalassemia status (thal), ST depression induced by exercise (oldpeak), and maxi-mum heart rate achieved (thalach). Elevated values of oldpeak generally increase the predicted probability of heart disease, whereas variations in thalach provide information regarding cardiac performance and exercise tolerance. The prominence of these attributes confirms that the model relies on clinically relevant risk factors rather than arbitrary statistical patterns.

An important advantage of SHAP analysis is its ability to indicate both the magnitude and direction of feature influence. Positive SHAP values increase the predicted risk of heart disease, while negative values reduce the predicted probability. This capability provides deeper insight into model behavior and enables validation of the learned relationships against established medical knowledge.

2) *Patient-Level Interpretation:* In addition to global ex-planations, SHAP was employed to generate patient-specific *interpretations* through force plots. These visualizations ex-plain how individual clinical features contribute to a particu-lar prediction outcome and allow healthcare professionals to understand the reasoning behind model decisions on a case-by-case basis.

The SHAP force plot shown in Fig. 6 illustrates how differ-ent clinical parameters either increase or decrease the predicted risk of heart disease for a specific patient. Features displayed with positive contributions push the prediction toward the

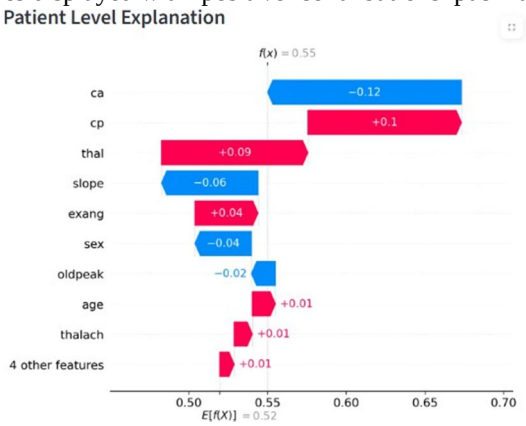


Fig. 6. SHAP force plot illustrating patient-level feature contributions toward heart disease prediction.

presence of heart disease, whereas features with negative con-tributions reduce the predicted risk. The cumulative effect of these contributions determines the final classification generated by the Random Forest model.

Patient-level explanations provide significant practical value in clinical environments. Rather than presenting only a binary prediction outcome, the proposed framework identifies the specific risk factors responsible for the prediction. This en-ables clinicians to validate model decisions, assess individual cardiovascular risk profiles, and communicate findings more effectively to patients. Overall, the SHAP analysis demonstrates that the proposed framework achieves a balance between predictive performance and interpretability. By providing both global feature impor-tance and individualized explanations, the system enhances transparency, strengthens clinician trust, and improves the practical applicability of machine learning for heart disease prediction and clinical decision support.

C. Discussion

The experimental findings demonstrate that the proposed Explainable AI-driven framework effectively combines pre-dictive performance with interpretability for heart disease pre-diction. The comparative evaluation revealed that the Random Forest classifier consistently outperformed Logistic Regression across all performance metrics, including accuracy, precision, recall, and F1-score. These results suggest that ensemble learn-ing methods are better suited for modeling the complex and often non-linear relationships that exist among cardiovascular risk factors.

The superior performance of Random Forest can be at-tributed to its ability to aggregate predictions from multiple decision trees, thereby reducing overfitting and improving generalization capability. Unlike Logistic Regression, which assumes a linear relationship between input features and the target variable, Random Forest can capture intricate feature interactions that are commonly observed in clinical datasets.

This characteristic enables the model to learn more repre-sentative decision boundaries and achieve higher predictive accuracy when classifying patients according to cardiovascular risk.

The ROC-AUC value of 0.857 further validates the effec-tiveness of the proposed framework. In medical prediction systems, the ability to distinguish between positive and nega-tive cases across varying decision thresholds is essential. The obtained AUC score indicates that the model possesses strong discriminative capability and can reliably separate patients with heart disease from those without the condition. Such performance highlights the suitability of the framework for supporting clinical screening and risk assessment tasks.

While predictive accuracy remains an important evaluation criterion, practical deployment of artificial intelligence in healthcare requires transparency and trust. One of the major limitations of many existing heart disease prediction models is their black-box nature, which restricts clinician understanding of prediction outcomes. To overcome this limitation, the proposed framework integrates SHAP-based explainability, en-abling both global and patient-specific interpretation of model decisions.

The SHAP analysis identified clinically meaningful predic-tors such as the number of major vessels (*ca*), chest pain type (*cp*), thalassemia status (*thal*), ST depression (*oldpeak*), and maximum heart rate achieved (*thalach*) as the most influential factors affecting prediction outcomes.

These findings are con-sistent with established cardiovascular risk indicators reported in medical literature, providing additional confidence in the reliability and validity of the model. Furthermore, patient-level explanations generated through SHAP force plots allow clinicians to understand the specific factors contributing to individual predictions, thereby improving transparency and facilitating informed decision-making.

Compared with conventional machine learning approaches that focus solely on predictive performance, the proposed framework offers a balanced combination of accuracy, inter-pretability, and clinical relevance. By integrating ensemble learning with explainable artificial intelligence, the system transforms machine learning predictions into actionable clin-ical insights. This capability enhances trust in AI-assisted diagnosis and supports the broader adoption of intelligent decision-support systems within healthcare environments.

Overall, the results demonstrate that combining Random Forest classification with SHAP-based explanation provides an effective and transparent approach for heart disease prediction. The framework not only achieves strong predictive perfor-mance but also addresses the growing need for explainability in healthcare artificial intelligence, thereby contributing to the development of trustworthy and clinically useful decision-support technologies.

V. CONCLUSION

This study presented an Explainable AI-driven framework for heart disease prediction using clinical parameters and ensemble learning techniques. The proposed approach utilized the UCI Heart Disease Dataset comprising 1,025 patient records and thirteen clinically relevant attributes to develop a predictive system capable of identifying the presence of heart disease. Two machine learning algorithms, Logistic Regression and Random Forest, were investigated and evaluated using standard classification metrics.

Experimental results demonstrated that the Random Forest classifier outperformed Logistic Regression across all eval-uation measures, including accuracy, precision, recall, and F1-score. The model achieved strong predictive performance while maintaining robust generalization capability. Further-more, ROC analysis confirmed the effectiveness of the frame-work, achieving an Area Under the Curve (AUC) value of 0.857, indicating reliable discrimination between patients with and without heart disease.

A key contribution of this research is the integration of SHapley Additive Explanations (SHAP) to enhance model transparency and interpretability. Through both global and patient-level explanations, the framework identified clinically significant predictors such as the number of major vessels (ca), chest pain type (cp), thalassemia status (thal), ST de-pression induced by exercise (oldpeak), and maximum heart rate achieved (thalach). These explanations provide valuable insights into the factors influencing prediction outcomes and improve trust in machine learning-assisted clinical decision-making.

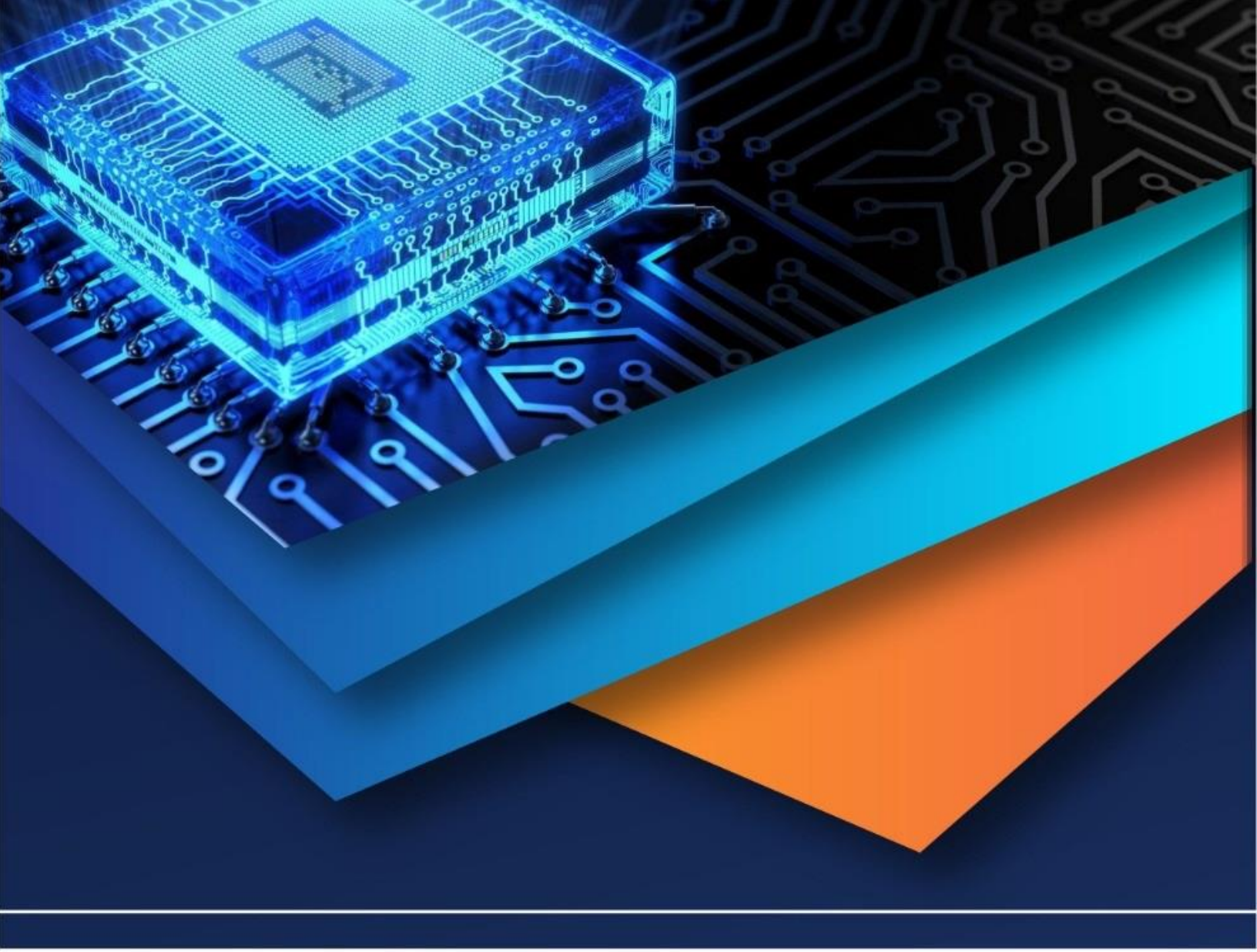
Unlike conventional black-box prediction systems, the pro-posed framework combines predictive accuracy with explain-ability, enabling healthcare professionals to understand the reasoning behind model decisions. This capability enhances transparency and supports the practical adoption of artifi-cial intelligence in clinical environments. By providing in-terpretable predictions and clinically meaningful explanations, the framework serves as a reliable decision-support tool for heart disease risk assessment.

Future work may focus on evaluating the framework using larger and more diverse clinical datasets, incorporating addi-tional machine learning and deep learning models, and imple-menting advanced validation strategies such as cross-validation and external dataset testing. Further improvements may also include real-time deployment within healthcare applications and the integration of additional explainability techniques to strengthen clinical usability and trust.

REFERENCES

[1] Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, School of Information and Computer Sciences, 2019.
[2] R. Detrano et al., “International application of a new probability al-gorithm for the diagnosis of coronary artery disease,” The American Journal of Cardiology, vol. 64, no. 5, pp. 304–310, 1989.
[3] M. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, “Heart Disease Dataset,” UCI Machine Learning Repository, 1988.
[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
[5] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[6] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[7] P. K. Anooj, “Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,” Journal of King Saud University – Computer and Information Sciences, vol. 24, no. 1, pp. 27–40, 2012.
[8] M. Akhil Jabbar, B. L. Deekshatulu, and P. Chandra, “Heart disease prediction system using associative classification and genetic algorithm,” Procedia Technology, vol. 10, pp. 183–192, 2013.
[9] K. Polat and S. Gu¨nes¸, “A hybrid approach to medical decision support system based on principal component analysis and adaptive neuro-fuzzy inference system,” Applied Mathematics and Computation, vol. 189, no. 2, pp. 1533–1544, 2007.
[10] M. Chicco and G. Jurman, “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone,” BMC Medical Informatics and Decision Making, vol. 20, no. 1, 2020.

[11] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.
[12] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, no. 1, pp. 56–67, 2020.
[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016, pp. 1135–1144.
[14] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018.
[15] D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) Program," AI Magazine, vol. 40, no. 2, pp. 44–58, 2019.
[16] A. Holzinger et al., "What do we need to build explainable AI systems for the medical domain?" arXiv preprint arXiv:1712.09923, 2017.
[17] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.
[18] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44–56, 2019.
[19] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.
[20] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
[21] J. Pearl, Causality: Models, Reasoning and Inference, 2nd ed. Cam-bridge, UK: Cambridge University Press, 2009.
[22] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY