Explainable AI-Driven Clinical Decision Support System for Cardiovascular Disease Prediction Using Hybrid Models

Romaan Khan[1], Muhammad Khan[1], Muhammad Younas[1], Nizam Ahmad[1], Abdullah[1], Islam Uddin[2*]

[1]Institute of Computer Science and Information Technology, University of Agriculture Peshawar, Pakistan

[2]Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan

*Correspondence: islamuddin@awkum.edu.pk

Citation | Khan. R, Khan. M, Younas. M, Ahmad. N, Abdullah, Uddin. I, "Explainable AI-Driven Clinical Decision Support System for Cardiovascular Disease Prediction Using Hybrid Models", IJIST, Vol. 8 Issue. 4 pp 1583-1603, July 2026

Received June 06, 2026 **Revised** | July 05, 2026 **Accepted** | July 09, 2026 **Published** | July 13, 2026

Artificial intelligence (AI) and machine learning have played increasingly important roles in healthcare, particularly in disease risk prediction, clinical data analysis, and decision support. Recent advances in ensemble learning and explainable AI (XAI) have created new opportunities to improve predictive performance and transparency in cardiovascular disease (CVD) risk assessment. However, conventional machine learning models may inadequately capture complex nonlinear relationships, while deep learning models often lack interpretability, limiting their reliability for clinical decision support. To address these limitations, this paper proposed an XAI-CDSS, an explainable ensemble learning framework for cardiovascular disease risk prediction and clinical decision support. First, cardiovascular clinical data are systematically preprocessed through data cleaning, normalization, categorical encoding, outlier handling, and class balancing. Secondly, Extreme Gradient Boosting (XGBoost) and a deep neural network (DNN) are employed as complementary learners, and their prediction probabilities are integrated through a weighted ensemble strategy to improve predictive robustness and generalization. Thirdly, Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) provide global feature-level and patient-specific explanations, respectively. Finally, the framework is evaluated on a benchmark CVD dataset comprising 1,025 patient records, including an independent test set of 205 records. Experimental results demonstrate that XAI-CDSS achieves an accuracy of 94.63%, precision of 96.94%, recall of 92.23%, and F1-score of 94.53%. Compared with the individual constituent models and existing CVD prediction approaches considered in this study, XAI-CDSS demonstrates competitive predictive performance while providing complementary global and patient-specific explainability, highlighting its potential for transparent and reliable clinical decision support.

Keywords: Cardiovascular Disease Risk Prediction; Explainable Artificial Intelligence (XAI); Clinical Decision Support System (CDSS); Ensemble Learning.

Introduction:

Cardiovascular disease (CVD) remains a major global health challenge, accounting for a substantial proportion of mortality worldwide and imposing a considerable burden on healthcare systems. Each year, cardiovascular conditions contribute to millions of deaths, many of which may be mitigated through timely risk assessment, early detection, and appropriate clinical intervention [1]. CVD risk is influenced by multiple interacting factors, including age, blood pressure, cholesterol levels, smoking, physical activity, and other physiological, genetic, and lifestyle characteristics [2]. Conventional clinical assessment largely relies on medical expertise and the interpretation of patient-specific clinical information, which can become increasingly complex as the volume and heterogeneity of healthcare data expand [3]. The growing availability of medical datasets and electronic health records (EHRs) has therefore increased interest in intelligent computational systems capable of supporting timely and data-driven clinical decisions [4]. In this context, artificial intelligence (AI) and machine learning (ML) have emerged as promising approaches for healthcare applications, particularly for disease risk assessment and prediction [5]. These techniques can analyze complex clinical data and identify nonlinear relationships and feature interactions that may be difficult to capture using conventional analytical approaches [6]. Consequently, clinical decision support systems (CDSSs) incorporating AI-driven predictive models have gained increasing attention as tools for assisting healthcare professionals in risk assessment and clinical decision-making [7]. However, the limited interpretability and transparency of many high-performing AI models remain important barriers to their trustworthy use in healthcare. Explainable artificial intelligence (XAI) seeks to address these limitations by meaningful explainability is essential for developing transparent and trustworthy clinical decision-support frameworks for CVD risk prediction [8].

Recent advances in machine learning (ML) and deep learning (DL) have significantly improved the potential of computational approaches for CVD prediction; however, several challenges continue to limit the development of reliable and interpretable clinical decision-support systems [9]. Recent advances in machine learning (ML) and deep learning (DL) have significantly improved the potential of computational approaches for CVD prediction; however, several challenges continue to limit the development of reliable and interpretable clinical decision-support systems [10]. Advanced DL models, such as Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs), can learn complex feature representations and nonlinear patterns, but their decision-making processes are often difficult to interpret [9]. This lack of transparency remains an important concern in healthcare settings, where interpretability, accountability, and confidence in model-supported decisions are essential [11]. Moreover, many existing studies primarily emphasize predictive performance without simultaneously addressing robustness, generalizability, and explainability within a unified framework [12]. Hybrid and ensemble learning strategies have been increasingly explored to leverage the complementary strengths of different algorithms; nevertheless, effectively integrating heterogeneous predictive learners with meaningful explainability remains an important challenge [13]. In addition, evaluations based on limited datasets and validation settings may restrict the generalizability of predictive models across broader patient populations and real-world clinical environments [14]. Therefore, there is a need for decision-support frameworks that combine complementary learning paradigms with transparent explanations of model predictions [15]. This requirement is particularly important in healthcare, where unreliable or insufficiently interpretable predictions may limit confidence in AI-assisted clinical decision-making [16]. Integrating ensemble learning with explainability techniques such as SHAP and LIME therefore represents a promising direction for developing transparent, reliable, and interpretable CVD risk-prediction and clinical decision-support systems [17].

Recent years have witnessed growing interest in cardiovascular disease (CVD) prediction, with increasing emphasis on improving predictive performance, robustness, and model interpretability through machine learning (ML), ensemble learning, and explainable artificial intelligence (XAI). Early studies commonly employed conventional classifiers, including Logistic Regression, Support Vector Machines (SVM), and Decision Trees; however, individual models may have limitations in capturing complex nonlinear relationships and heterogeneous interactions within clinical data. Consequently, ensemble learning has gained considerable attention by combining multiple predictive models to improve classification robustness and generalization. For example, hybrid ensemble approaches integrating Random Forest, SVM, and XGBoost have demonstrated improved predictive stability compared with individual models in heart disease prediction [18]. Similarly, explainable ensemble learning approaches have been investigated to improve CVD classification while maintaining interpretable predictions for clinical applications [19]. Soft-voting ensemble strategies have further demonstrated the potential of combining multiple learning algorithms with explainability mechanisms to improve predictive performance and generalization [20]. For structured clinical data, optimized XGBoost models combined with feature-selection strategies have also shown strong predictive capability in heart disease classification [21]. Furthermore, weighted ensemble approaches integrating XGBoost with deep learning models have been explored to leverage their complementary learning capabilities for cardiovascular risk prediction [22]. In parallel, explainable AI techniques, particularly Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), have been increasingly employed to provide global and patient-specific insights into model predictions and improve the transparency of AI-assisted clinical decision-making [23]. Despite these advances, existing studies often emphasize either predictive optimization through individual or ensemble models or the post-hoc interpretation of model outputs. The systematic integration of complementary ensemble learners with both global and patient-specific explainability within a unified clinical decision-support framework remains comparatively underexplored [24]. This gap motivates the development of an integrated framework that jointly considers predictive performance, complementary model learning, and multi-level explainability for CVD risk prediction.

Building on these observations, an important research gap remains in developing a unified framework that effectively integrates complementary ensemble learning with global and patient-specific explainability for cardiovascular disease (CVD) risk prediction and clinical decision support. To address this gap, this study proposes XAI-CDSS, an explainable ensemble learning framework for cardiovascular disease risk prediction and clinical decision support. The proposed framework integrates Extreme Gradient Boosting (XGBoost) and a Deep Neural Network (DNN) as complementary predictive learners, with their prediction probabilities combined through a weighted averaging strategy to improve predictive robustness and generalization [25]. This integration is designed to leverage the complementary capabilities of tree-based and deep learning approaches for capturing complex feature interactions and nonlinear patterns in clinical data. To enhance prediction transparency, XAI-CDSS incorporates Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) as complementary explainability mechanisms. SHAP provides global insights into the relative contributions of clinical features across the predictive model, whereas LIME generates patient-specific explanations that clarify the factors influencing individual predictions. The predictive performance of the proposed framework is evaluated using established classification measures, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion-matrix analysis. XAI-CDSS integrates complementary ensemble prediction and multi-level explainability within a unified framework to provide reliable,

transparent, and interpretable computational support for CVD risk prediction and clinical decision-making.

The main contributions of this paper are summarized as follows:

An XAI-CDSS ensemble learning framework integrating XGBoost and DNN to capture structured feature interactions and complex nonlinear patterns for CVD risk prediction.

A weighted probability-level fusion strategy combining XGBoost and DNN predictions to improve predictive robustness and generalization.

A multi-level explainability approach integrating SHAP for global feature interpretation and LIME for patient-specific explanations.

A unified clinical decision-support framework combining ensemble prediction and explainability for transparent and reliable CVD risk assessment.

The remainder of this paper is organized as follows. Section 2 presents the proposed XAI-CDSS architecture and methodology, including data preprocessing, ensemble learning, and explainability mechanisms. Section 3 describes the experimental setup and presents the performance evaluation, comparative analysis, explainability results, and discussion. Section 4 concludes the study, summarizes its limitations, and outlines directions for future research.

Materials and Methods:

This section presents the methodology used to develop XAI-CDSS, an explainable ensemble learning framework for cardiovascular disease (CVD) risk prediction and clinical decision support. The proposed methodology integrates data preprocessing, complementary predictive learning, ensemble-based probability fusion, and explainable artificial intelligence within a unified framework. The methodology consists of four main stages: dataset preparation, data preprocessing and feature engineering, ensemble prediction, and model explainability. The clinical data are first prepared and preprocessed to ensure their suitability for model development. XGBoost and a DNN are then employed as complementary predictive learners, and their probability outputs are combined to generate the final CVD risk prediction. SHAP and LIME are subsequently applied to provide global and patient-specific explanations, respectively. The following subsections describe each component of the proposed methodology in detail.

Framework Architecture:

The proposed XAI-CDSS is an explainable ensemble learning framework designed to provide accurate, robust, and transparent cardiovascular disease (CVD) risk prediction for clinical decision support. The framework integrates complementary predictive learning with explainable artificial intelligence (XAI) to improve both predictive performance and model interpretability. It consists of four sequential stages: dataset acquisition, data preprocessing and feature engineering, ensemble prediction, and model explainability [26]. The framework begins by acquiring a benchmark CVD dataset containing demographic, clinical, and lifestyle-related attributes. The collected data are subsequently processed through data cleaning, categorical feature encoding, normalization, and class balancing to improve data quality and ensure suitability for predictive modeling. After preprocessing, two complementary predictive learners, XGBoost and a DNN, are independently trained to capture structured feature interactions and complex nonlinear relationships within the clinical data. The prediction probabilities generated by both models are then integrated using a weighted probability-level ensemble strategy to produce the final CVD risk prediction. To improve the transparency of the predictive process, the framework incorporates two complementary explainability techniques. Shapley Additive Explanations (SHAP) provide global feature importance by quantifying the contribution of each clinical variable across the entire dataset, whereas Local Interpretable Model-Agnostic Explanations (LIME) generate patient-specific explanations that identify the factors influencing individual predictions. The integration of ensemble

learning with multi-level explainability enables XAI-CDSS to provide reliable, interpretable, and transparent support for CVD risk prediction and clinical decision-making. The overall architecture and workflow of the proposed XAI-CDSS framework are illustrated in Figure 1.

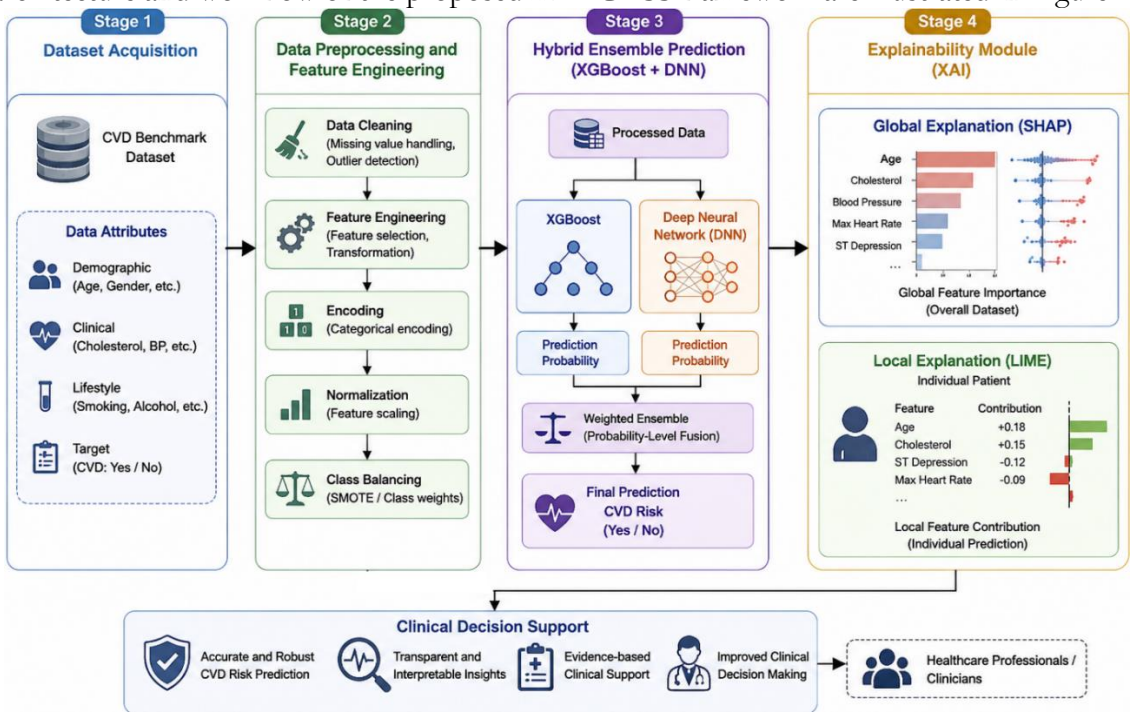

Figure 1. The architecture of the proposed XAI-CDSS framework

Benchmark Dataset:

The proposed XAI-CDSS framework was developed and evaluated using the publicly available Heart Disease Dataset, originally obtained from the UCI Machine Learning Repository and made available through Kaggle for research purposes. This benchmark dataset has been widely used for developing and evaluating machine learning (ML), deep learning (DL), ensemble learning, and explainable artificial intelligence approaches for cardiovascular disease (CVD) prediction. More recently, [22] employed a cardiovascular disease dataset in their weighted ensemble and explainable machine-learning study, further demonstrating the continued relevance of benchmark cardiovascular datasets for evaluating predictive and interpretable AI approaches. The dataset used in the present study comprises 1,025 records, 13 predictor variables, and one binary target variable indicating the presence or absence of CVD.

The predictor variables represent demographic, clinical, laboratory, and physiological characteristics associated with cardiovascular risk, including age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, ST-segment slope, number of major vessels, and thalassemia status. A detailed description of the variables is provided in Table 1.

Table 1. Description of Variables Used in the CVD Dataset

Predictor Variable	Explanation
Variable	Description
Age	Represents the age of the patient in years.
Sex	Indicates biological sex of the patient, where male and female are encoded numerically.

Chest Pain Type (cp)	Represents the category of chest pain experienced by the patient, which serves as an important indicator of cardiovascular abnormalities.
Resting Blood Pressure (trestbps)	Measures resting systolic blood pressure recorded in mmHg.
Serum Cholesterol (chol)	Represents cholesterol concentration measured in mg/dL.
Fasting Blood Sugar (fbs)	Indicates whether fasting blood sugar exceeds the clinically accepted threshold.
Resting Electrocardiographic Results (restecg)	Reflects ECG abnormalities observed during rest.
Maximum Heart Rate Achieved (thalach)	Measures peak heart rate during exercise testing.
Exercise-Induced Angina (exang)	Indicates whether chest pain occurs during physical activity.
ST Depression (oldpeak)	Measures depression induced by exercise relative to resting condition.
Slope	Represents the slope of the peak exercise ST segment.
Number of Major Vessels (ca)	Indicates the number of major blood vessels identified through fluoroscopy.
Thalassemia (thal)	Represents blood disorder-related diagnostic findings.

The dataset was selected because of its established use in CVD prediction research, clinically relevant predictor variables, relatively balanced class distribution, and suitability for comparison with related ML, DL, ensemble learning, and explainable AI approaches. Its relevance is further supported by recent cardiovascular prediction research, including the ensemble and explainability-based study of [22]. However, because cardiovascular datasets and experimental protocols may differ across studies, cross-study performance comparisons should be interpreted cautiously. The complete dataset contained 499 negative (Class 0) and 526 positive (Class 1) records, indicating a relatively balanced class distribution. The distribution of samples in the training and independent test sets is presented in Table 2.

Table 2. Dataset Partitioning and Class Distribution for Model Development and Evaluation

Dataset Split	Total Records	Negative (Class 0)	Positive (Class 1)	Percentage (%)
Total Dataset	1,025	499	526	100
Training Set	820	397	423	80
Independent Test Set	205	102	103	20

Before model development, the dataset was partitioned into 80% training data and 20% independent test data using stratified sampling to preserve the class proportions of the original dataset. As summarized in Table 2, the training set contained 820 records, including 397 Class 0 and 423 Class 1 samples, whereas the independent test set contained 205 records, including 102 Class 0 and 103 Class 1 samples. The comparable class proportions across the two subsets preserved the overall class distribution. The training set was used exclusively for model development, preprocessing-dependent operations, optimization, and internal validation, whereas the independent test set was held out from model development and reserved exclusively for final performance evaluation. This separation was maintained to reduce the risk of information leakage and to provide a more reliable assessment of the predictive performance of the proposed XAI-CDSS on previously unseen samples.

Data Preprocessing:

Data preprocessing is a fundamental step in developing reliable machine learning and deep learning models because the quality of the input data directly influences predictive performance, model stability, and generalization. Clinical datasets may contain duplicate observations, heterogeneous feature scales, categorical variables, and potential outliers, all of which may adversely affect the learning process and reduce the reliability of model predictions [27]. Therefore, an effective preprocessing pipeline is essential to improve data quality, preserve clinically meaningful information, and generate representative inputs for predictive modeling. In this study, a systematic preprocessing framework was implemented before model development [28]. The preprocessing pipeline consists of five sequential stages: duplicate record inspection, feature normalization, outlier detection, categorical feature encoding, and class balancing. These procedures were designed to improve data consistency, reduce potential sources of bias, enhance model convergence, and prepare the dataset for the proposed XAI-CDSS framework.

Clinical datasets may contain duplicate observations, heterogeneous feature scales, categorical variables, and potential outliers that can affect model training and predictive reliability. Therefore, a systematic preprocessing pipeline was implemented to improve data consistency and prepare the cardiovascular disease (CVD) data for the proposed XAI-CDSS framework [29]. First, duplicate observations were identified to reduce data redundancy and minimize potential bias during model development. Numerical features with different measurement scales were subsequently transformed to a common range of [0, 1] using Min–Max normalization:

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

Where x represents the original feature value, x_min and x_max denote the minimum and maximum values of the corresponding feature, and x_norm represents the normalized value. Potential outliers in continuous clinical variables were identified using the interquartile range (IQR) criterion:

$$IQR = Q_3 - Q_1 \quad (2a)$$

$$LB = Q_1 - 1.5 \times IQR \quad (2b)$$

$$UB = Q_3 + 1.5 \times IQR \quad (2c)$$

Where Q_1, and Q_3, represent the first and third quartiles, respectively, while LB and UB, denote the lower and upper bounds. Observations outside these bounds were identified as potential outliers and examined during preprocessing. The feature distributions and potential outliers identified using the IQR criterion are presented in Figure 2.

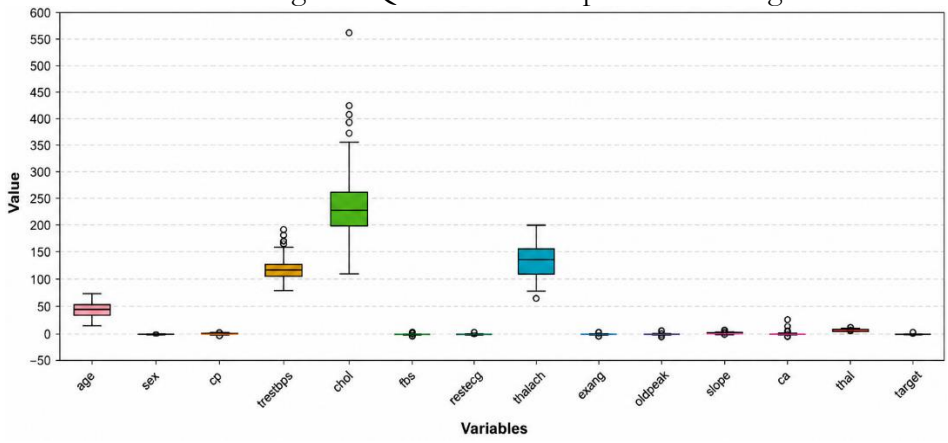

Figure 2. Distribution of Clinical Variables and Outlier Detection

Categorical clinical variables were then transformed into numerical representations to ensure compatibility with the XGBoost and DNN learners. The preprocessing procedure

preserved the clinical meaning of categorical attributes while converting them into machine-readable representations suitable for predictive modeling. Class distribution was also examined to assess potential imbalance between the two outcome classes. The original dataset contained 499 Class 0 (absence of CVD) and 526 Class 1 (presence of CVD) observations [30].

The Synthetic Minority Oversampling Technique (SMOTE) was considered to balance the minority class by generating synthetic observations through interpolation between neighboring minority-class samples:

$$x_{\text{new}} = x_i + \lambda(x_k - x_i) \quad (3)$$

Where x_i represents a minority-class observation, x_k denotes a neighboring minority-class sample, and λ is a randomly generated value within [0,1]. For model development, oversampling was restricted to the training data to prevent synthetic information from entering the independent test set and causing data leakage. Figure 3 illustrates the class distribution of the complete dataset for descriptive purposes, showing 499 Class 0 and 526 Class 1 observations before balancing and the corresponding balanced distribution of 526 observations per class after SMOTE-based oversampling [31].

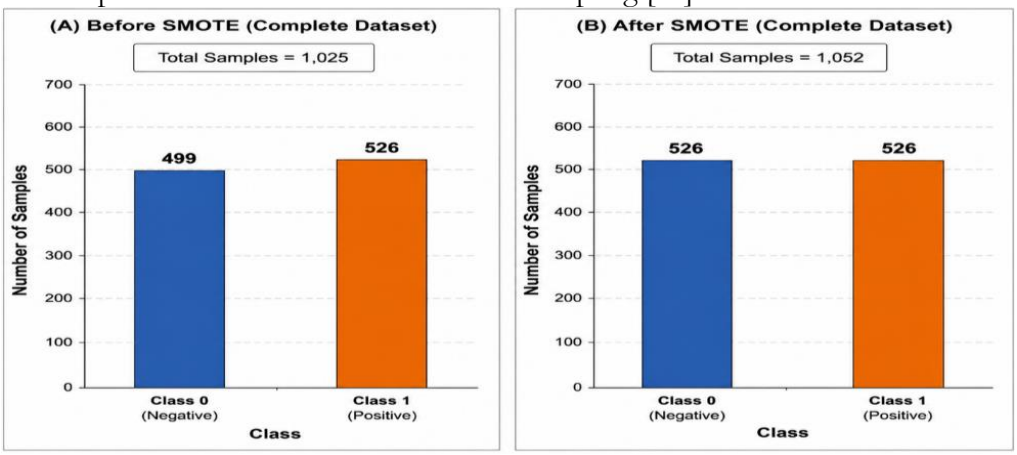

Figure 3. Class Distribution of the CVD Dataset Before and After SMOTE-Based Balancing

Feature Selection:

Despite the high accuracy of ML and DL models in healthcare prediction tasks, a large number of these models are black-box systems. For this reason, they are not widely used by doctors and have been used in actual clinical applications only to a limited degree [32]. To overcome this challenge, explainability mechanisms were incorporated into the proposed CDSS. The explainability module is based on two complementary techniques such as SHAP and LIME. Both methods can be integrated to give global-level and patient-level interpretability, enhancing transparency and physician confidence.

SHAP:

SHAP is a game-theory-based explainability approach which measures the impact of each clinical feature individually on the model's prediction. The proposed framework was designed to include SHAP to identify the most significant cardiovascular characteristics impacting prediction [33]. Global explanations will evaluate the importance of the features of the entire set, and be used to identify important cardiovascular features that can be used for disease prediction. For instance, some factors like type of chest pain, cholesterol level, maximum heart rate, ST depression can be the most important cardiovascular risk factors.

Local explanations are explanations for a particular patient of why the prediction is made. For example, a patient's cardiovascular risk may be greatly increased by his cholesterol level and electrocardiographic abnormalities. The SHAP explainability mechanism helps

enhance transparency by providing healthcare professionals with insights into model decisions and clinically relevant risk factors.

For a feature i, the Shapley value is defined as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (4)$$

Where F denotes the complete set of input features, S is a subset of features that excludes feature i, and $f(S)$ represents the model output obtained using the features contained in subset S. The term $f(S \cup \{i\}) - f(S)$ measures the marginal contribution of feature i, while the weighting factor accounts for all possible feature combinations. Therefore, ϕ_i represents the average contribution of feature $i$to the model output across different feature coalitions.

For an individual patient x, the additive SHAP explanation is expressed as:

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j \quad (5)$$

Where ϕ_0 is the expected model output over the reference dataset, M is the total number of input features, and ϕ_j represents the SHAP contribution of the jth feature for the specific patient. A positive SHAP value shifts the model output toward the CVD class, whereas a negative value shifts it toward the non-CVD class. The magnitude $|\phi_j|$indicates the strength of the feature's contribution.

The incorporation of XAI techniques is one of the proposed XAI-CDSS's major contributions, ensuring transparency and clinical trustworthiness. The proposed system is globally and locally interpretable using SHAP and LIME, which are different from the traditional black-box ML models. The SHAP analysis gives a global feature importance, by estimating the contribution of each feature over the whole dataset. The findings confirm that some of the factors most closely related with CVD are chest pain type, cholesterol levels, resting blood pressure, and maximum heart rate. This global interpretability is useful for understanding the overall decision-making behaviour of the model, which is of value to the clinician. Figure 4 shown the summary of the SHAP.

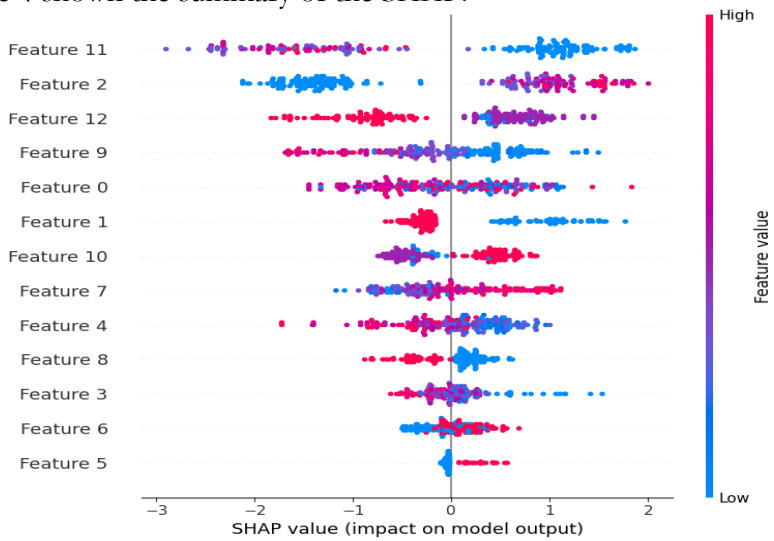

Figure 4. SHAP Summary Plot for Global Feature Contribution Analysis

LIME:

The proposed framework was complemented by LIME to generate patient-specific explanations. LIME is not as comprehensive as SHAP, as it only offers explanations for individual predictions. It produces meaningful explanations by simulating the prediction behaviour of the black-box model on a particular patient record. There are several stages to

the LIME explanation process. The first step is to choose a patient case for analysis. Second, several instances of the sample are created around the chosen one. Third, the prediction behavior of the model is observed. Lastly, an interpretable local explanation is generated to uncover what features were positively or negatively associated with the prediction. The LIME explanations are particularly useful in medical settings where doctors need to provide reasoning for each patient before adopting AI recommendations. The framework might, for instance, indicate that high cholesterol, exercise-induced angina and elevated blood pressure were major factors in the increased risk of CVD for a specific patient. LIME generates a local explanation by finding a simple interpretable model g that best approximates the black-box model f in the neighbourhood of the patient instance x_i. The LIME objective function is defined as:

$$\xi(xi) = argmin\,[\,L(f, g, \pi xi) + \Omega(g)] \quad (6)$$

Where $\xi(x_i)$ is the local explanation for patient x_i, g is a simple interpretable model (e.g., a linear model), f is the original, black-box model, L is a fidelity loss function that quantifies how well g approximates f in the neighborhood around πx_i is a locality kernel that defines the neighborhood around x_i and $\Omega(g)$ is a complexity penalty term that encourages simplicity and interpretability of the model for the explanation. Figure 5 shows the LIME explanation for individual patient prediction.

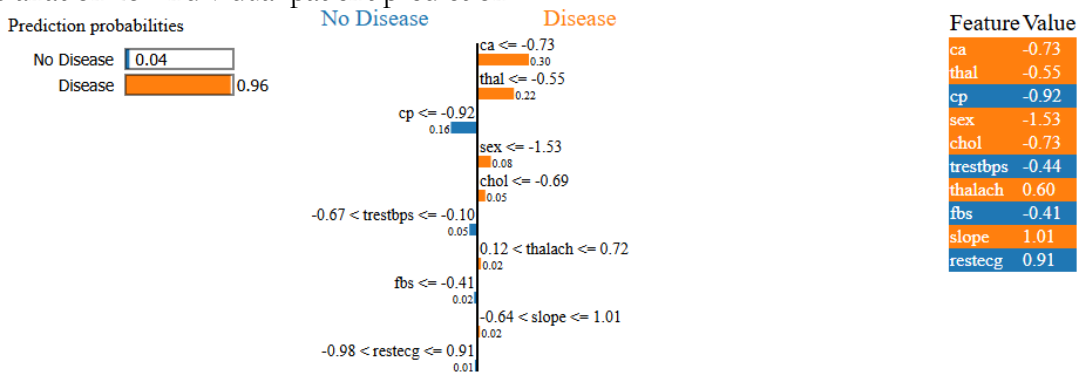

Figure 5. LIME-Based Patient-Specific Explanation for CVD Risk Prediction

Feature Relevance Analysis:

Feature relevance analysis was performed to examine the relationships among the clinical variables and assess their potential contribution to CVD prediction. This analysis is important because highly redundant or weakly informative variables may increase model complexity without providing meaningful predictive information. The proposed framework therefore examines both inter-feature relationships and model-based feature importance to obtain a clearer understanding of the information represented by the clinical predictors. Correlation analysis was used to quantify the relationships among the input variables and identify potential dependencies or redundancy within the feature space [34]. The resulting correlation matrix provides an overview of the direction and strength of pairwise associations among the clinical variables. Model-based feature importance was subsequently examined to assess the relative contribution of individual predictors to CVD prediction. These analyses support interpretation of the feature space and provide complementary information regarding the relevance of clinical predictors. The correlation structure among the clinical variables is presented in Figure 6

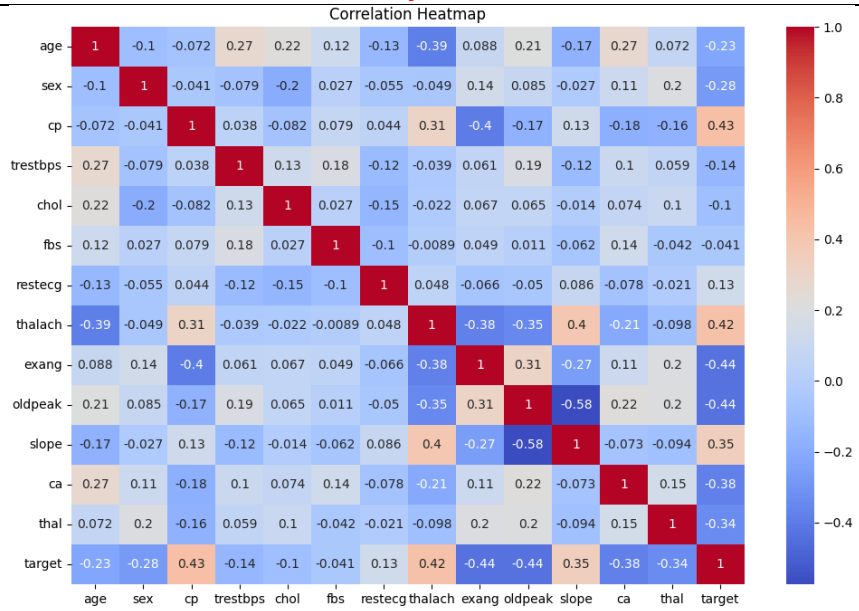

Figure 6. Correlation Heatmap of Clinical Features in the CVD Dataset

Ensemble Prediction Model:

The predictive component of XAI-CDSS integrates XGBoost and a DNN as complementary learners for cardiovascular disease (CVD) risk prediction. XGBoost is employed to capture structured feature interactions and discriminative patterns within tabular clinical data, whereas the DNN learns complex nonlinear relationships through multiple hidden layers [35]. The complementary learning characteristics of these models provide the basis for integrating their predictive outputs within a unified ensemble framework.

The proposed architecture consists of two parallel prediction branches. The XGBoost branch generates a CVD probability using an ensemble of sequentially optimized decision trees, while the DNN branch independently estimates the probability through nonlinear transformations of the clinical features. The probability outputs of both learners are subsequently combined at the probability level using a weighted averaging strategy. The ensemble probability is formulated as:

$$P_{\text{ens}}(x) = w_X P_X(x) + w_D P_D(x) \quad (7)$$

Subject to:

$$w_X + w_D = 1,\ w_X, w_D \geq 0 \quad (8)$$

Where $P_X(x)$and $P_D(x)$represent the predicted CVD probabilities generated by XGBoost and DNN, respectively, while $w_X$and $w_D$denote their corresponding ensemble weights. The final class prediction is determined using a decision threshold τ:

$$\hat{y} = \begin{cases} 1, & P_{\text{ens}}(x) \geq \tau \\ 0, & P_{\text{ens}}(x) < \tau \end{cases} \quad (9)$$

Where $\hat{y} = 1$, indicates the predicted presence of CVD and $\hat{y} = 0$, indicates its predicted absence. The ensemble design integrates the complementary predictive characteristics of tree-based and deep learning models rather than relying on a single learning paradigm.

The proposed Clinical Decision Support System (CDSS) is based on a hybrid prediction model that combines XGBoost with DNN to enhance the accuracy and robustness of the prediction of cardiovascular diseases. The proposed hybrid prediction model is shown in Figure 7. The rationale behind merging these two predictive models is that they learn complementary representations of the data. The Deep Neural Network can learn nonlinear patterns and hierarchical feature representations with a number of hidden layers, while

XGBoost can learn the structured relationships and complex interactions between features from the tabular clinical data. The proposed approach combines both models into a single framework in order to benefit from the strengths of each of the models and, at the same time, to reduce the limitations of each one.

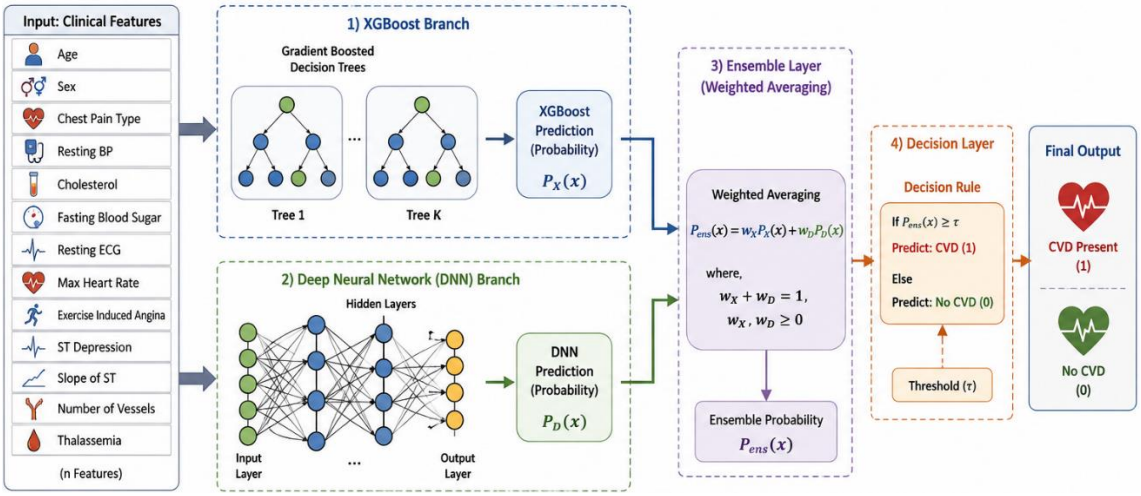

Figure 7. Hybrid prediction model integrating XGBoost and Deep Neural Network for cardiovascular disease prediction

Performance Evaluation Metrics:

The predictive performance of the proposed XAI-CDSS framework was evaluated using multiple classification metrics to provide a comprehensive assessment of its ability to distinguish between CVD and non-CVD cases. The evaluation metrics included accuracy, precision, recall (sensitivity), specificity, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). Accuracy measures overall classification correctness, whereas precision quantifies the reliability of positive predictions. Recall evaluates the ability to correctly identify patients with CVD, while specificity measures the correct identification of non-CVD cases. The F1-score provides a balanced measure of precision and recall, and ROC-AUC evaluates the overall discriminative ability of the model across different classification thresholds.

Accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Precision is calculated as:

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Recall, also referred to as sensitivity, is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

Specificity is calculated as:

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

The F1-score represents the harmonic mean of precision and recall:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (14)$$

Where TP and TN denote correctly classified CVD and non-CVD cases, respectively, while FP and FN represent incorrectly classified positive and negative cases.The ROC curve evaluates model discrimination by plotting the true positive rate (TPR) against the false positive rate (FPR) across classification thresholds:

$$TPR = \frac{TP}{TP+FN}, \; FPR = \frac{FP}{FP+TN} \quad (15)$$

The area under the ROC curve is expressed as:

$$AUC = \int_0^1 T\,PR(FPR)\,d(FPR) \quad (16)$$

A higher AUC indicates stronger overall discrimination between CVD and non-CVD cases. Collectively, these metrics provide complementary measures of classification correctness, positive-case detection, negative-case discrimination, and overall predictive performance.

Experimental Results and Analysis:

This section presents the experimental evaluation and analysis of the proposed XAI-CDSS framework for cardiovascular disease (CVD) risk prediction. The framework integrates XGBoost and DNN as complementary predictive learners through a weighted ensemble strategy and incorporates SHAP and LIME for global and patient-specific explainability. The final predictive performance was evaluated on an independent test set of 205 records using accuracy, precision, recall, specificity, F1-score, ROC-AUC, and confusion-matrix analysis. The experimental evaluation focuses on two main objectives: (i) assessing the predictive performance of XAI-CDSS for distinguishing between CVD and non-CVD cases and (ii) examining the interpretability of model predictions through global and patient-specific explanations.

Experimental Environment:

The proposed XAI-CDSS framework was implemented and evaluated in a controlled computational environment to ensure consistent model development and performance assessment. The implementation was carried out using Python 3.10, with Scikit-learn, XGBoost, and TensorFlow supporting data preprocessing, model development, training, and evaluation. SHAP and LIME were employed to generate global and patient-specific explanations of the model predictions, respectively. The predictive framework consisted of XGBoost and a DNN as complementary learners. Both models were developed using the same preprocessed clinical data, and their prediction probabilities were integrated through a weighted probability-level ensemble strategy to generate the final CVD risk prediction. The dataset was divided into 80% training and 20% independent test sets, with 205 records reserved exclusively for final evaluation. The independent test set was kept separate from model training and optimization to provide an unbiased assessment of predictive performance. The performance of XAI-CDSS was evaluated using accuracy, precision, recall, specificity, F1-score, ROC-AUC, and confusion-matrix analysis. SHAP and LIME analyses were further used to assess global feature contributions and patient-specific prediction behavior. The hardware, software, model, data, and evaluation configurations used in the experimental framework are summarized in Table 3.

Table 3. Experimental Setup Configuration

Category	Component	Specification
Hardware	Processor (CPU)	Standard multi-core processor
	Graphics Processor (GPU)	GPU-enabled computing environment
	Memory (RAM)	Standard system memory
Software	Operating System	Windows/Linux-based environment
	Programming Language	Python 3.10
	ML/DL Libraries	Scikit-learn, XGBoost, TensorFlow
	Explainability Libraries	SHAP, LIME
Model Configuration	Predictive Learners	XGBoost and Deep Neural Network (DNN)
	Ensemble Strategy	Weighted probability-level averaging
Data Configuration	Dataset Split	80% training and 20% independent testing
	Independent Test Set	205 records
Evaluation	Performance Metrics	Accuracy, Precision, Recall, Specificity, F1-score, ROC-AUC

	Additional Analysis	Confusion matrix and explainability analysis

Hyperparameter Configuration:

Hyperparameter configuration plays an important role in controlling model complexity, learning behavior, and generalization performance. The XGBoost and DNN components of the proposed XAI-CDSS require different parameter settings because of their distinct learning mechanisms. Therefore, model-specific hyperparameters were configured during model development using the training data, while the independent test set was reserved exclusively for final performance evaluation. For XGBoost, the main configuration parameters included the number of estimators, learning rate, maximum tree depth, minimum child weight, subsampling ratio, and feature-sampling ratio. These parameters control the boosting process, tree complexity, and degree of sampling during model training. For the DNN, the configuration included the network architecture, activation functions, optimizer, learning rate, batch size, number of epochs, validation split, and early stopping. The network consisted of an input layer corresponding to the 13 clinical predictors, followed by two hidden layers with 64 and 32 neurons, respectively, and an output neuron with sigmoid activation for binary CVD risk prediction. ReLU activation was used in the hidden layers, while the model was optimized using Adam with binary cross-entropy loss. The hyperparameter configuration used for model development is summarized in Table 4.

Table 4. Hyperparameter Configuration of the XGBoost and DNN Predictive Learners

Model	Hyperparameter	Value
XGBoost	Number of Estimators (n_estimators)	100
	Learning Rate (learning_rate)	0.10
	Maximum Tree Depth (max_depth)	6
	Minimum Child Weight (min_child_weight)	1
	Subsample Ratio (subsample)	0.80
	Column Sampling (colsample_bytree)	0.80
	Objective Function	Binary Logistic
	Evaluation Metric	Log Loss
	Random State	42
DNN	Input Layer	13 neurons
	Hidden Layer 1	64 neurons (ReLU)
	Hidden Layer 2	32 neurons (ReLU)
	Output Layer	1 neuron (Sigmoid)
	Optimizer	Adam
	Learning Rate	0.001
	Loss Function	Binary Cross-Entropy
	Batch Size	32
	Maximum Epochs	100
	Validation Split	20%
	Early Stopping	Enabled

Ensemble Model Performance Evaluation:

The predictive performance of the proposed XAI-CDSS was evaluated using the independent test set of 205 patient records. The ensemble model achieved an overall classification accuracy of 94.63%, demonstrating strong performance in distinguishing between CVD and non-CVD cases. The model achieved a precision of 96.94%, recall (sensitivity) of 92.23%, specificity of 97.06%, and F1-score of 94.53%, indicating balanced predictive performance across multiple evaluation criteria. The proposed ensemble integrates XGBoost and DNN as complementary predictive learners. XGBoost captures discriminative patterns and interactions among structured clinical features, whereas the DNN learns complex

nonlinear relationships within the clinical data. Their prediction probabilities are integrated through the weighted probability-level ensemble strategy described in Section 2.5. This integration enables both learning mechanisms to contribute to the final CVD risk prediction, rather than relying on a single predictive model. The performance of XGBoost, DNN, and the proposed XAI-CDSS ensemble was further compared to assess the contribution of ensemble integration. The comparative results demonstrate the relative performance of the individual learners and their combined ensemble configuration, as presented in Figure 8.

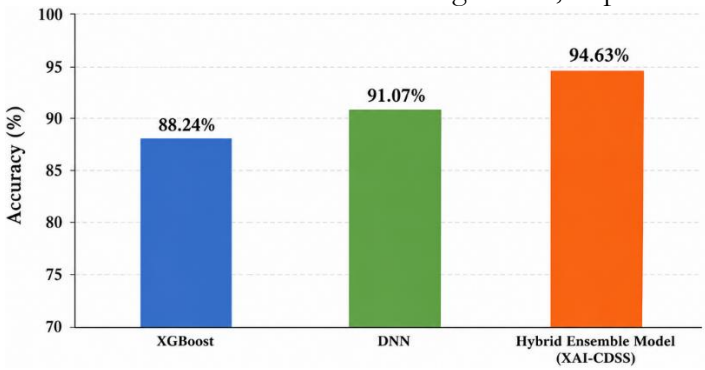

Figure 8. Comparative Accuracy Performance of the Proposed XAI-CDSS and Individual Predictive Models

Confusion Matrix and Diagnostic Performance Analysis:

The classification performance of the proposed XAI-CDSS was further examined using confusion-matrix analysis to assess its ability to distinguish between CVD and non-CVD cases. Among the 205 records in the independent test set, the model correctly classified 95 CVD cases as true positives (TP) and 99 non-CVD cases as true negatives (TN). In contrast, 3 non-CVD cases were incorrectly classified as positive (FP), while 8 CVD cases were incorrectly classified as negative (FN). Based on these outcomes, XAI-CDSS achieved a precision of 96.94%, sensitivity of 92.23%, specificity of 97.06%, and F1-score of 94.53%, together with an overall accuracy of 94.63%. The high specificity indicates strong discrimination of non-CVD cases, whereas the sensitivity demonstrates the model's ability to identify most CVD-positive cases. However, the presence of eight false-negative predictions remains clinically important because missed positive cases may require particular attention when considering the framework for decision-support applications. These findings highlight the importance of evaluating both positive- and negative-class performance rather than relying solely on overall accuracy. The distribution of correct and incorrect predictions is presented in Figure 9.

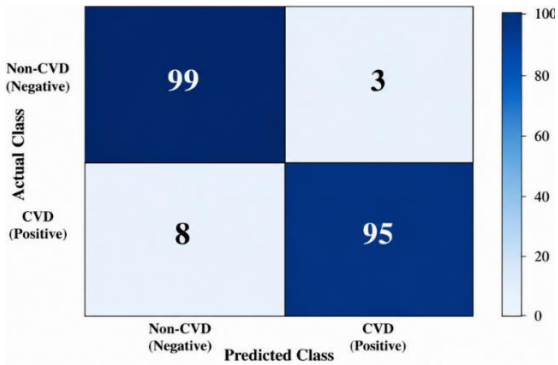

Figure 9. Confusion Matrix Analysis of XAI-CDSS for CVD Classification

Receiver Operating Characteristic (ROC) Curve and Discriminative Capability Analysis:

The discriminative capability of the proposed XAI-CDSS was evaluated using Receiver Operating Characteristic (ROC) curve analysis. The ROC curve represents the relationship

between the true positive rate (sensitivity) and false positive rate (1 − specificity) across different classification thresholds, providing a threshold-independent assessment of the ability of each model to distinguish between CVD and non-CVD cases. The Area Under the ROC Curve (AUC) was used as the primary measure of discrimination, with a higher AUC indicating stronger overall class-separation capability. The proposed XAI-CDSS achieved the highest AUC of 0.953, demonstrating superior discriminative performance among the evaluated models. In comparison, the DNN achieved an AUC of 0.937, followed by XGBoost with 0.913, Random Forest with 0.892, and SVM with 0.874. XAI-CDSS therefore improved the AUC by 0.016 over DNN, 0.040 over XGBoost, 0.061 over Random Forest, and 0.079 over SVM. These comparative results indicate that integrating the complementary predictions of XGBoost and DNN through the proposed ensemble strategy provides stronger discrimination than the evaluated individual models. The ROC analysis complements the threshold-dependent classification metrics reported in the preceding sections by evaluating predictive discrimination across a range of decision thresholds. All evaluated models achieved AUC values substantially above the 0.500 chance level, while XAI-CDSS demonstrated the strongest overall discriminative capability. The comparative ROC curves and corresponding AUC values are presented in Figure 10.

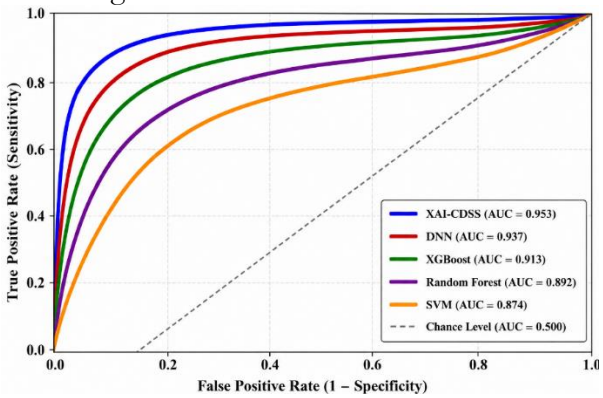

Figure 10. Comparative ROC Curve and AUC Analysis of XAI-CDSS and Baseline Models

Comparison with State-of-the-Art Methods:

The proposed XAI-CDSS was compared with existing machine learning and ensemble-based approaches for cardiovascular disease (CVD) prediction to evaluate its relative predictive performance. The selected studies employ different learning strategies, including XGBoost, optimized boosting models, soft-voting classifiers, and weighted ensemble methods, with several approaches incorporating SHAP and LIME for model interpretation. The proposed XAI-CDSS integrates XGBoost and DNN through weighted probability-level fusion and incorporates SHAP and LIME to provide global and patient-specific explanations. Table 5 presents the comparative results using accuracy (ACC), specificity (SP), sensitivity (SN), F1-score, Matthews correlation coefficient (MCC), and the area under the ROC curve (AUC). XAI-CDSS achieved an accuracy of 94.63%, specificity of 97.06%, sensitivity of 92.23%, F1-score of 94.53%, MCC of approximately 0.895, and AUC of 0.953. These results demonstrate strong and balanced predictive performance across multiple evaluation measures. The comparison also highlights the integration of complementary ensemble learning and multi-level explainability within the proposed framework. However, direct numerical comparisons should be interpreted cautiously because the referenced studies differ in dataset characteristics, sample sizes, preprocessing procedures, validation protocols, and experimental settings.

Table 5. Performance Comparison of the Proposed XAI-CDSS with Existing CVD Prediction Methods

Studies	Accuracy (%)

[26]	71.13
[18]	88.00
[21]	92.40
[19]	93.00
[20]	89.10
[22]	93.00
Proposed XAI-CDSS	94.63

Discussion:

The experimental results demonstrate that the proposed XAI-CDSS achieved strong predictive and discriminative performance for CVD risk prediction. On the independent test set, the framework obtained an accuracy of 94.63%, precision of 96.94%, sensitivity of 92.23%, specificity of 97.06%, F1-score of 94.53%, MCC of 0.895, and AUC of 0.953. The confusion matrix showed 95 true positives, 99 true negatives, 3 false positives, and 8 false negatives. The higher specificity than sensitivity indicates stronger identification of non-CVD cases, whereas the false-negative cases highlight an important limitation because missed positive cases remain particularly relevant in clinical risk assessment. The observed performance can be interpreted in relation to the complementary learning characteristics of XGBoost and DNN. XGBoost is well suited to structured tabular data and can capture nonlinear feature interactions through gradient-boosted decision trees, whereas the DNN learns distributed nonlinear representations through successive hidden layers. Their probability-level integration allows both learners to contribute to the final prediction rather than relying exclusively on a single modeling paradigm. The comparative ROC analysis supports this complementarity: XAI-CDSS achieved an AUC of 0.953, compared with 0.937 for DNN and 0.913 for XGBoost individually, as well as 0.892 for Random Forest and 0.874 for SVM. These findings suggest that ensemble integration provided additional discriminative value in the present experimental setting. However, the improvement cannot be attributed solely to the ensemble architecture without further ablation, repeated cross-validation, and statistical significance testing.

The findings also extend recent research on ensemble-based and explainable CVD prediction. [20] investigated interpretable machine-learning approaches with SHAP and LIME, while [21] combined ensemble learning with explainable AI for heart-disease prediction. [22] further demonstrated the integration of soft-voting classification with SHAP and LIME, and Hutagalung and Andrianingsih [23] explored optimized tree-based models with SHAP explanations. Hasnat et al. [24] investigated weighted ensemble learning with explainability on a larger cardiovascular dataset, whereas [26] demonstrated explainable XGBoost-based cardiovascular risk assessment in patients with Type 2 Diabetes Mellitus. XAI-CDSS follows this broader progression from individual predictive models toward ensemble and explainable systems, while specifically integrating XGBoost–DNN probability-level ensemble learning with complementary SHAP and LIME explanations. Numerical comparisons across these studies should nevertheless be interpreted cautiously because their datasets, populations, preprocessing procedures, validation protocols, and evaluation metrics differ substantially. The explainability analysis provides an additional perspective beyond predictive performance. SHAP characterizes how individual features contribute to model outputs across the analyzed samples, enabling examination of overall feature-attribution patterns and the direction and magnitude of their contributions. Features such as chest pain type, cholesterol, resting blood pressure, and maximum heart rate showed notable contributions in the model's explanations. These findings indicate which variables the trained model relied upon most strongly; however, high SHAP importance should not be interpreted as evidence of causal clinical importance. Correlated predictors may also influence attribution

patterns, and the resulting explanations remain dependent on the trained model and underlying dataset.

LIME complements this global perspective by approximating model behavior within the local neighborhood of an individual prediction. Patient-specific explanations can therefore reveal which variables support or oppose a particular predicted CVD outcome. The complementary use of SHAP and LIME is important because agreement between global and local explanations is not guaranteed: a feature that is influential across the dataset may have limited influence for a particular patient, while another feature may strongly affect an individual prediction without being globally dominant. This distinction provides a more informative interpretation of model behavior than relying on a single feature-importance ranking. Nevertheless, SHAP and LIME remain post-hoc explanation methods; their outputs explain model behavior rather than clinical causality, and their stability, consistency, and clinical meaningfulness require further validation by healthcare professionals.

Several limitations affect the interpretation of these findings. The study used a single benchmark dataset containing 1,025 records, and the independent test set originated from the same underlying dataset as the development data. Consequently, generalizability to different populations, institutions, and clinical environments has not yet been established. The current experiments also do not include external multicenter validation, prospective evaluation, formal calibration analysis, subgroup fairness assessment, or statistical significance testing of performance differences. Future studies should therefore evaluate XAI-CDSS on larger and geographically diverse external cohorts, examine calibration and subgroup performance, conduct repeated or nested cross-validation, statistically assess improvements over baseline models, and evaluate the stability and clinical relevance of SHAP and LIME explanations with domain experts. Such validation is necessary before the framework can be considered for real-world clinical decision-support deployment.

Conclusion:

Cardiovascular disease risk prediction requires computational approaches that provide reliable predictive performance while maintaining sufficient transparency for clinical decision support. This study proposed XAI-CDSS, an explainable ensemble learning framework for cardiovascular disease risk prediction and clinical decision support. The framework integrates XGBoost and DNN as complementary predictive learners to capture structured feature interactions and complex nonlinear patterns in clinical data. A weighted probability-level fusion strategy combines the outputs of both learners to generate the final CVD risk prediction. The framework further incorporates SHAP and LIME to provide complementary levels of explainability. SHAP enables global analysis of feature contributions, whereas LIME provides patient-specific explanations for individual predictions. This integration establishes a unified framework that combines ensemble-based prediction with global and local interpretability. The experimental results demonstrate that XAI-CDSS achieved an accuracy of 94.63%, precision of 96.94%, sensitivity of 92.23%, specificity of 97.06%, F1-score of 94.53%, MCC of 0.895, and AUC of 0.953 on the independent test set. The ROC comparison further showed stronger discriminative performance than the evaluated individual baseline models, supporting the potential benefit of complementary ensemble integration in the present experimental setting. However, the proposed framework was evaluated using a single benchmark dataset, which limits conclusions regarding generalizability across diverse populations and healthcare environments. Future research will focus on external and multicenter validation, larger and more diverse cohorts, model calibration, statistical significance analysis, subgroup evaluation, and prospective clinical assessment. Further evaluation of SHAP and LIME explanation stability and validation by clinical experts will also be necessary before considering XAI-CDSS for real-world clinical decision-support applications.

Declarations:

Ethics Approval and Consent to Participate: Not applicable. This study used a publicly available benchmark dataset and did not involve direct recruitment of human participants or collection of new patient data.

Consent for Publication: Not applicable.

Availability of Data and Materials: The dataset analyzed in this study is publicly available through the UCI Machine Learning Repository and its publicly distributed Kaggle version. No new clinical data were collected as part of this study.

Competing Interests: The authors declare that they have no competing interests.

Funding: This research received no external funding.

Authors' Contributions: R.K., M.K., M.Y., N.A., A.U and I.U. contributed to the conception, methodology, analysis, interpretation of results, and preparation of the manuscript. All authors reviewed and approved the final version of the manuscript.

Acknowledgments: The authors acknowledge the institutional support provided by the Institute of Computer Science and Information Technology, University of Agriculture Peshawar, Pakistan, and the Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan.

References:

[1] O. Gaidai, Y. Cao, and S. Loginov, "Global Cardiovascular Diseases Death Rate Prediction," *Curr. Probl. Cardiol.*, vol. 48, no. 5, May 2023, doi: 10.1016/j.cpcardiol.2023.101622.

[2] V. Vaccarino and J. D. Bremner, "Stress and cardiovascular disease: an update," *Nat. Rev. Cardiol.*, vol. 21, no. 9, pp. 603–616, Sep. 2024, doi: 10.1038/S41569-024-01024-Y.

[3] N. Kalra, P. Verma, and S. Verma, "Advancements in AI based healthcare techniques with FOCUS ON diagnostic techniques," *Comput. Biol. Med.*, vol. 179, p. 108917, Sep. 2024, doi: 10.1016/J.COMPBIOMED.2024.108917.

[4] J. Ferdush, M. Begum, and S. T. Hossain, "ChatGPT and Clinical Decision Support: Scope, Application, and Limitations," *Ann. Biomed. Eng.*, vol. 52, no. 5, pp. 1119–1124, May 2024, doi: 10.1007/S10439-023-03329-4.

[5] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, "Machine learning and artificial intelligence in research and healthcare," *Injury*, vol. 54, pp. S69–S73, May 2023, doi: 10.1016/J.INJURY.2022.01.046.

[6] N. R. & H. A. H. Mohammed Badawy, "Healthcare predictive analytics using machine learning and deep learning techniques: a survey," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 40, 2023, doi: https://doi.org/10.1186/s43067-023-00108-y.

[7] "(PDF) Transforming Healthcare with AI and Machine Learning: Revolutionizing Patient Care Through Advanced Analytics." Accessed: Jul. 25, 2026. [Online]. Available: https://www.researchgate.net/publication/387303877_Transforming_Healthcare_with_AI_and_Machine_Learning_Revolutionizing_Patient_Care_Through_Advanced_Analytics

[8] Roy Okonkwo, Adebola Folorunso, Foyeke Ogundipe, and Clement Yayra Tettey, "Explainable Artificial Intelligence (Al) through human-AI collaborative frameworks: Quantifying trust and interpretability in high-stakes decisions," *Comput. Sci. IT Res. J.*, vol. 6, no. 5, pp. 333–354, Jun. 2025, doi: 10.51594/CSITRJ.V6I5.1934.

[9] I. Uddin *et al.*, "A hybrid residue based sequential encoding mechanism with XGBoost improved ensemble model for identifying 5-hydroxymethylcytosine modifications," *Sci. Reports 2024 141*, vol. 14, no. 1, pp. 20819-, Sep. 2024, doi: 10.1038/s41598-024-71568-z.

[10] I. Uddin, N. Iqbal, and S. A. C. Bukhari, "Discriminative Sequence Features for DNA-Binding Protein Detection Using ML and DL Techniques," *Proc. - 2025 IEEE Int. Conf. Futur. Mach. Learn. Data Sci. FMLDS 2025*, pp. 416–421, 2025, doi: 10.1109/FMLDS67896.2025.00078.

[11] S. C. E. Nouis, V. Uren, and S. Jariwala, "Evaluating accountability, transparency, and bias in AI-assisted healthcare decision- making: a qualitative study of healthcare professionals' perspectives in the UK," *BMC Med. Ethics*, vol. 26, no. 1, Dec. 2025, doi: 10.1186/S12910-025-01243-Z.

[12] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," *Artif. Intell. Rev. 2021 555*, vol. 55, no. 5, pp. 3503–3568, Nov. 2021, doi: 10.1007/S10462-021-10088-Y.

[13] S. Dhanka, A. Sharma, A. Kumar, S. Maini, and H. Vundavilli, "Advancements in Hybrid Machine Learning Models for Biomedical Disease Classification Using Integration of Hyperparameter-Tuning and Feature Selection Methodologies: A Comprehensive Review," *Arch. Comput. Methods Eng. 2025 331*, vol. 33, no. 1, pp. 289–324, Jun. 2025, doi: 10.1007/S11831-025-10309-5.

[14] Y. Huang, J. Li, M. Li, and R. R. Aparasu, "Application of machine learning in predicting survival outcomes involving real-world data: a scoping review," *BMC Med. Res. Methodol.*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/S12874-023-02078-1.

[15] F. Valente, S. Paredes, J. Henriques, T. Rocha, P. de Carvalho, and J. Morais, "Interpretability, personalization and reliability of a machine learning based clinical decision support system," *Data Min. Knowl. Discov. 2022 363*, vol. 36, no. 3, pp. 1140–1173, Apr. 2022, doi: 10.1007/S10618-022-00821-8.

[16] A. Leiva-Araos, C. Contreras, H. Kaushal, and Z. Prodanoff, "Predictive Optimization of Patient No-Show Management in Primary Healthcare Using Machine Learning," *J. Med. Syst.*, vol. 49, no. 1, Dec. 2025, doi: 10.1007/S10916-025-02143-W.

[17] Adewale Abayomi Adeniran, Amaka Peace Onebunne, and Paul William, "Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making," *World J. Adv. Res. Rev.*, vol. 23, no. 3, pp. 2447–2658, Sep. 2024, doi: 10.30574/WJARR.2024.23.3.2936.

[18] P. S. Asih, Y. Azhar, G. W. Wicaksono, and D. R. Akbi, "Interpretable Machine Learning Model For Heart Disease Prediction," *Procedia Comput. Sci.*, vol. 227, pp. 439–445, Jan. 2023, doi: 10.1016/J.PROCS.2023.10.544.

[19] A. Adekoya, F. Saeed, W. Ghaban, and S. N. Qasem, "Ensemble learning approach with explainable AI for improved heart disease prediction," *Front. Pharmacol.*, vol. 16, p. 1654681, Dec. 2025, doi: 10.3389/FPHAR.2025.1654681/TEXT.

[20] S. Walia, S. Walia, A. Bhardwaj, S. Arora, S. Aggarwal, and P. Siwach, "Explainable soft-voting classifier for heart disease prediction using SHAP and LIME," *Discov. Comput. 2026 291*, vol. 29, no. 1, pp. 83-, Feb. 2026, doi: 10.1007/S10791-026-09972-4.

[21] P. H. Hutagalung and A. Andrianingsih, "Heart Disease Classification Using Optimised XGBoost and Random Forest with SHAP Explanations," *sinkron*, vol. 10, no. 1, pp. 330–342, Jan. 2026, doi: 10.33395/SINKRON.V10I1.15544.

[22] M. A. Hasnat, M. Jobayer, M. M. H. Shawon, and M. G. R. Alam, "Interpretable Heart Disease Prediction via a Weighted Ensemble Model: A Large-Scale Study with SHAP and Surrogate Decision Trees," Nov. 2025, Accessed: Jul. 25, 2026. [Online]. Available: https://arxiv.org/pdf/2511.01947

[23] R. Agrawal, T. Gupta, S. Gupta, S. Chauhan, P. Patel, and S. Hamdare, "Fostering trust and interpretability: integrating explainable AI (XAI) with machine learning for enhanced disease prediction and decision transparency," *Diagnostic Pathol. 2025 201*, vol. 20, no. 1, pp. 105-, Sep. 2025, doi: 10.1186/S13000-025-01686-3.

[24] S. Khan, I. Uddin, F. K. Alarfaj, and N. Almusallam, "A novel computational framework for tumor-specific T cell antigen identification using a deep neural network," *J. Comput. Mol. Des. 2026 401*, vol. 40, no. 1, pp. 150-, Jun. 2026, doi: 10.1007/S10822-026-00861-Y.

[25] M. Alzamel, I. Uddin, S. Khan, N. Dilshad, and N. Ahmad, "Deep learning-based identification of N6-methyladenine sites via hybrid feature fusion and SHAP-driven feature selection," *Sci. Reports 2026 161*, vol. 16, no. 1, pp. 22385-, May 2026, doi:

10.1038/s41598-026-50667-z.

[26] M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, "An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus," Sep. 2020, Accessed: Jul. 25, 2026. [Online]. Available: https://arxiv.org/pdf/2009.06629

[27] I. Uddin *et al.*, "Deep-m6Am: a deep learning model for identifying N6, 2'-O-Dimethyladenosine (m6Am) sites using hybrid features," *AIMS Bioeng. 2025 1145*, vol. 12, no. 1, pp. 145–161, 2025, doi: 10.3934/BIOENG.2025006.

[28] P. Shah, M. Shukla, N. H. Dholakia, and H. Gupta, "Predicting cardiovascular risk with hybrid ensemble learning and explainable AI," *Sci. Reports 2025 151*, vol. 15, no. 1, pp. 17927-, May 2025, doi: 10.1038/s41598-025-01650-7.

[29] A. Ur Rahman, N. Iqbal, I. Uddin, and S. A. C. Bukhari, "Multifaceted Fusion of Glove and Additive Attention Models for Scalable Fake News Detection," *Proc. - 2025 IEEE Int. Conf. Futur. Mach. Learn. Data Sci. FMLDS 2025*, pp. 363–368, 2025, doi: 10.1109/FMLDS67896.2025.00069.

[30] S. Anbazhagan, S. S. Ranganathan, M. Alagarsamy, and R. Kuppusamy, "A comprehensive hybrid model for early detection of cardiovascular diseases using integrated CardioXGBoost and long short-term memory networks," *Biomed. Signal Process. Control*, vol. 95, p. 106281, Sep. 2024, doi: 10.1016/J.BSPC.2024.106281.

[31] M. A. Talukder, A. S. Talaat, M. Kazi, and A. Khraisat, "XAI-HD: an explainable artificial intelligence framework for heart disease detection," *Artif. Intell. Rev. 2025 5812*, vol. 58, no. 12, pp. 385-, Oct. 2025, doi: 10.1007/S10462-025-11385-6.

[32] J. Boobalan and L. Kiruthika, "Explainable AI for CVD Risk Prediction: An Ensemble Learning Approach with Clinical Interpretability," *Proc. 3rd Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2025*, pp. 1036–1043, 2025, doi: 10.1109/ICSCDS65426.2025.11167113.

[33] Masroor Shah, Fazal Malik, Muhammad Suliman, Noor Rahman, Irfan Ullah, Sana Ullah, Romaan Khan, & Salman Alam, "Dark Data in Accident Prediction: Using AdaBoost and Random Forest for Improved Accuracy," *J. Comput. Biomed. Informatics*, vol. 7, no. 2, 2024, [Online]. Available: https://jcbi.org/index.php/Main/article/view/531

[34] I. Uddin, S. Noor, Y. A. Ali, S. Khan, and M. Al-Razgan, "A hybrid deep learning framework for accurate N6,2'-O-Dimethyladenosine site prediction," *Biophys. Chem.*, vol. 335, p. 107639, Aug. 2026, doi: 10.1016/J.BPC.2026.107639.

[35] S. Khan, I. Uddin, S. Noor, S. A. AlQahtani, and N. Ahmad, "N6-methyladenine identification using deep learning and discriminative feature integration," *BMC Med. Genomics 2025 181*, vol. 18, no. 1, pp. 58-, Mar. 2025, doi: 10.1186/S12920-025-02131-6.