



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IX **Month of publication:** September 2026

DOI: <https://doi.org/10.22214/ijraset.2026.84833>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Adaptive Hybrid Cardiovascular Disease Prediction Using Deep Neural Networks and Anomaly-Aware Risk Scoring

Golagani Usha Sri¹, Dr. Golagani Lavanya Devi²

Department of Computer Science and System Engineering, Andhra University, Visakhapatnam, Andhra Pradesh

Abstract: Cardiovascular disease (CVD) is a major health concern, making early risk identification an important task in healthcare decision support. This paper presents an adaptive hybrid cardiovascular disease prediction framework that combines a Deep Neural Network (DNN), Isolation Forest-based anomaly detection, and SHAP (SHapley Additive exPlanations) for interpretable risk assessment. The system uses clinical attributes from the UCI Cleveland Heart Disease dataset. Categorical features are transformed using one-hot encoding and the resulting features are scaled using StandardScaler. The DNN learns nonlinear relationships among clinical attributes and estimates the probability of cardiovascular disease, while the Isolation Forest identifies unusual patient patterns. These outputs are integrated using a hybrid risk-scoring mechanism in which the disease probability contributes 80% and the normalized anomaly component contributes 20%. SHAP is used to identify the features that influence individual predictions. The complete framework is implemented as an interactive Streamlit dashboard that presents disease probability, anomaly score, hybrid risk score, risk category, and feature-level explanations. An automated PDF reporting module additionally summarizes the prediction and clinical interpretation. The proposed framework therefore integrates prediction, anomaly awareness, explainability, and automated reporting into a unified cardiovascular disease risk-assessment system.

Keywords: Cardiovascular disease prediction, deep neural networks, anomaly detection, Isolation Forest, hybrid risk scoring, SHAP, explainable artificial intelligence, Streamlit.

I. INTRODUCTION

Cardiovascular diseases are among the major health challenges worldwide. Early identification of individuals who may be at increased risk can support timely clinical evaluation and appropriate intervention. The availability of clinical datasets and advances in artificial intelligence have created opportunities to develop computational systems capable of identifying complex relationships among cardiovascular risk factors. Traditional machine learning approaches have been widely applied to cardiovascular disease prediction using clinical variables. However, a prediction system that only provides a disease classification may provide limited information about unusual patient patterns or the reasons behind an individual prediction.

Deep Neural Networks can model nonlinear relationships among multiple clinical variables and can therefore be used to estimate disease probability. However, a DNN-based prediction alone does not explicitly identify whether a patient record contains an unusual combination of characteristics. To address this aspect, the implemented system incorporates an Isolation Forest for anomaly detection. Another important requirement in healthcare-oriented artificial intelligence is interpretability. Black-box predictions can be difficult to understand and evaluate. Therefore, SHAP Explainable AI is incorporated into the proposed framework to identify the contribution of individual features to the prediction.

The implemented system combines these components into a single workflow. Clinical data are preprocessed using categorical encoding and feature scaling, followed by DNN-based disease probability estimation and Isolation Forest-based anomaly detection. The two outputs are combined into a hybrid risk score. SHAP is then used to explain the prediction, and the results are presented through a Streamlit dashboard and downloadable PDF report.

A. Background and Motivation

Cardiovascular disease is a major healthcare concern, and early risk identification can support timely clinical assessment. Clinical attributes such as age, sex, chest pain, blood pressure, cholesterol, maximum heart rate, and other diagnostic factors provide useful information for cardiovascular risk prediction.

The motivation of this work is to develop a prediction system that goes beyond simple disease classification by combining Deep Neural Network (DNN) prediction, Isolation Forest anomaly detection, and SHAP-based Explainable AI. The framework provides an anomaly-aware hybrid risk score and explains the important features influencing the prediction through an interactive dashboard and automated report.

B. Problem Statement

Existing cardiovascular disease prediction systems mainly focus on disease classification and may not identify unusual patient patterns or clearly explain individual predictions. The lack of anomaly-aware risk assessment and model interpretability can limit the usefulness of AI-based prediction systems. Therefore, this work aims to develop an integrated framework that combines Deep Neural Network-based disease prediction, Isolation Forest anomaly detection, and SHAP Explainable AI to provide an interpretable hybrid cardiovascular risk assessment.

II. RELATED WORK

Krittanawong et al. [1] conducted a systematic review and meta-analysis of machine-learning applications in cardiovascular disease prediction. The study analyzed 103 cohorts from 55 studies involving approximately 3.38 million individuals and evaluated machine-learning approaches for coronary artery disease, stroke, heart failure, and cardiac arrhythmias. The results demonstrated promising predictive performance, with boosting and custom-built algorithms achieving pooled AUC values of 0.88 and 0.93, respectively, for coronary artery disease, while SVM, boosting, and CNN approaches also demonstrated strong performance for stroke prediction. However, the authors highlighted the importance of appropriate algorithm selection, validation, standardized evaluation, and clinical assessment before deployment. This study motivates the development of more comprehensive cardiovascular prediction frameworks that combine predictive modeling with additional mechanisms for anomaly detection and interpretability.

Rahim et al. [2] proposed an integrated machine learning framework, termed MaLCaDD, for cardiovascular disease prediction. The framework performs missing-value handling, class balancing using SMOTE, feature selection using feature importance, and prediction using an ensemble of Logistic Regression and K-Nearest Neighbor classifiers. The framework was evaluated on Framingham, Heart Disease, and Cleveland datasets, achieving reported accuracies of 99.1%, 98.0%, and 95.5%, respectively. The study demonstrates the effectiveness of integrated machine-learning approaches for cardiovascular disease prediction. However, the framework primarily focuses on predictive classification and does not incorporate the anomaly-aware risk scoring and SHAP-based individual prediction explanations implemented in the proposed system.

Ghosh et al. [3] proposed an efficient cardiovascular disease prediction framework using feature-selection and hybrid machine-learning techniques. The study used a combined dataset comprising Cleveland, Long Beach VA, Switzerland, Hungarian, and Statlog datasets. Relief and LASSO methods were employed for feature selection, followed by hybrid classifiers based on bagging and boosting techniques, including Random Forest Bagging, KNN Bagging, AdaBoost, and Gradient Boosting. The Random Forest Bagging Method with Relief feature selection achieved a reported accuracy of 99.05%. However, the framework primarily focuses on feature selection and predictive classification. In contrast, the proposed system integrates Deep Neural Network-based prediction with Isolation Forest anomaly detection, hybrid risk scoring, and SHAP-based explainability to provide a more comprehensive risk-assessment workflow.

Li et al. [4] developed an automated cardiovascular disease prediction framework using machine learning and electronic medical record data. The study analyzed 101,110 unique patient records and compared models using cross-sectional features with models combining cross-sectional and longitudinal clinical features. The Random Forest model using combined cross-sectional and longitudinal features achieved the best reported AUC of 0.902. The study also showed that the machine-learning models could reduce the proportion of patients requiring screening compared with the ACC/AHA Pooled Cohort Equations. However, the study primarily focused on population-specific ASCVD prediction using EMR data. In contrast, the proposed system uses a Deep Neural Network with Isolation Forest-based anomaly detection, hybrid risk scoring, SHAP-based explainability, and an interactive reporting interface for individual cardiovascular risk assessment.

Rahmathullah et al. [5] proposed a deep learning-based framework for cardiovascular disease prediction and risk assessment that integrates clinical, genetic, imaging, and sequential electronic health record data. The framework employs CNNs for medical image analysis, RNN/LSTM models for sequential data, and attention mechanisms for multimodal integration. SHAP and Grad-CAM are incorporated to improve model interpretability, while transfer learning and federated learning are considered to improve generalizability and data privacy. The study demonstrates the potential of deep learning for comprehensive cardiovascular risk assessment.

However, the framework primarily focuses on multimodal deep learning, whereas the proposed system uses structured clinical data and integrates a DNN with Isolation Forest-based anomaly detection, hybrid risk scoring, SHAP explanations, and an interactive reporting interface.

III. RESEARCH GAP

Existing cardiovascular prediction approaches primarily focus on supervised disease prediction. Such systems may not explicitly consider unusual patient patterns. In addition, complex deep learning models can be difficult to interpret without dedicated explainability techniques. The implemented project addresses these aspects by integrating DNN-based disease probability estimation, Isolation Forest-based anomaly detection, hybrid risk scoring, SHAP-based explainability, an interactive dashboard, and automated PDF report generation. The novelty is primarily in the integration of these components into a unified workflow rather than claiming that the individual algorithms themselves are novel.

Feature	Current System / Existing Work	Suggested System / Proposed Work
Prediction Approach	Primarily uses conventional machine learning, ensemble learning, or deep learning models for cardiovascular disease prediction.	Uses a Deep Neural Network (DNN) as the main disease prediction model.
Learning Framework	Most systems focus mainly on one model or a combination of supervised ML classifiers.	Combines supervised deep learning and unsupervised anomaly detection in a hybrid framework.
Risk Assessment	Generally provides disease classification, prediction probability, or conventional risk estimation.	Generates an anomaly-aware Hybrid Risk Score by combining DNN prediction and anomaly information.
Feature Processing	Uses preprocessing and, in some studies, feature selection techniques such as Relief and LASSO; EMR-based work also uses cross-sectional and longitudinal features.	Uses missing-value handling, one-hot encoding, and StandardScaler-based feature scaling for the project dataset.
Explainability	The reviewed approaches emphasize predictive performance and feature selection, but explainability is not a central integrated output across these systems.	Uses SHAP Explainable AI to show which features increase or reduce the predicted risk.
Clinical Decision Support	Mainly supports prediction or population-level risk assessment.	Provides an integrated decision-support workflow with prediction, anomaly analysis, explanation, risk classification, and reporting.

A. Research Objectives

The primary objective is to develop an interpretable hybrid cardiovascular disease prediction framework. The specific objectives are to:

- 1) Preprocess structured clinical data for machine-readable analysis.
- 2) Estimate disease probability using a trained DNN.
- 3) Analyze unusual patient patterns using Isolation Forest.
- 4) Combine the outputs into an anomaly-aware hybrid risk score.
- 5) Explain individual predictions using SHAP.
- 6) Present the results through an interactive dashboard and downloadable PDF report.

IV. PROPOSED METHODOLOGY

A. Dataset

The implementation uses a UCI heart disease dataset containing structured clinical attributes. The project dataset used during implementation contains 920 records and 16 original columns, including an identifier and dataset-source field. The clinical attributes used for prediction include age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting ECG, maximum heart rate, exercise-induced angina, ST depression, slope, number of major vessels, and thalassemia. The target information is represented by the disease-status field available in the dataset. The implementation documentation should use the exact dataset version and target transformation employed during model training.

B. Data Preprocessing

Before prediction, the input data are converted into the feature representation expected by the trained models. Missing values are handled during preprocessing. Categorical variables are converted using one-hot encoding, and numerical values are scaled using a StandardScaler. The final application uses 18 model input features stored in encoder_columns.pkl, ensuring that the feature order supplied to the scaler and models remains consistent.

One-hot encoding converts categories such as sex, chest pain type, resting ECG, exercise-induced angina, slope, and thalassemia into binary numerical features. The transformed feature vector is then aligned to the saved encoder-column order and scaled using the saved scaler object before it is passed to the DNN and Isolation Forest.

C. Deep Neural Network

The preprocessed feature vector is supplied to the trained Deep Neural Network (DNN) to estimate the probability of cardiovascular disease. The deployed application loads the final serialized model (dnn_model.h5) and applies the same feature alignment and StandardScaler transformation used during training. The exact layer-by-layer DNN architecture, including hidden layers, neurons, activation functions, dropout configuration, optimizer, learning rate, batch size, epoch count, validation strategy, and original train/test split ratio, is not recoverable from the final inference artifacts currently available for this paper. Therefore, these values are intentionally not invented. The final training script should be archived and used to populate this subsection before a submission requiring complete training reproducibility.

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 18)	0
dense_4 (Dense)	(None, 16)	304
latent space (Dense)	(None, 8)	136
dense_5 (Dense)	(None, 16)	144

D. Training Configuration and Evaluation Protocol

The final evaluation artifacts confirm that model performance was measured on a held-out test set containing 184 samples, represented by a confusion matrix with 67 true negatives, 15 false positives, 12 false negatives, and 90 true positives. Aggregate evaluation metrics were computed directly from these observed counts. The ROC curve reports an area under the curve (ROC-AUC) of 0.913. Because the exact original training code is not available in the current artifact set, the train/test split ratio, random seed, cross-validation procedure, decision threshold, and complete training hyperparameter configuration cannot be stated with certainty and are not fabricated here.

E. Isolation Forest Anomaly Analysis

The same scaled patient feature vector is analyzed using an Isolation Forest. The purpose of this component is not to classify disease directly, but to identify unusual or atypical patterns in the patient record relative to the feature distribution learned by the anomaly model. The implementation obtains an anomaly-related decision score and transforms it into a normalized component for use in the hybrid scoring stage.

F. Anomaly Normalization and Hybrid Risk Scoring

Let s denote the Isolation Forest decision score. The implementation converts this score into a normalized anomaly component using a sigmoid-based transformation:

$$A_{\text{norm}} = 1 / (1 + \exp(s))$$

The hybrid risk score is then computed as:

$$H = 0.80 P_{\text{DNN}} + 0.20 A_{\text{norm}}$$

where H is the hybrid risk score, P_{DNN} is the DNN disease probability, and A_{norm} is the normalized anomaly component. The DNN probability receives the larger weight because it is the primary supervised prediction output, while anomaly information acts as supplementary evidence. The 80:20 weighting is fixed in the current implementation and should be treated as an empirical design choice rather than an optimized clinical weighting.

G. Risk Classification

Hybrid Score	Risk Category
0.00–0.30	Low Risk
0.30–0.70	Moderate Risk
0.70–1.00	High Risk

H. SHAP Explainability

SHAP is incorporated to explain individual model predictions. The SHAP values indicate how each encoded feature contributes to moving the model output upward or downward for the analyzed input. The application maps encoded feature names into human-readable labels such as Thalassemia (Normal), Gender (Male), ST Segment Slope (Flat), Exercise-Induced Angina, Cholesterol, and Number of Major Vessels. A positive SHAP contribution indicates that a feature pushed the model output upward for the specific prediction, while a negative contribution pushed it downward. These values describe the behavior of the trained model for the analyzed input and should not automatically be interpreted as universal clinical causation.

V. SYSTEM ARCHITECTURE

The implemented workflow begins with clinical patient input and proceeds through preprocessing, one-hot encoding and feature scaling. The processed feature vector is supplied to the DNN for disease probability estimation and to the Isolation Forest for anomaly detection. The outputs are integrated by the Hybrid Risk Engine, followed by risk classification and SHAP-based explanation. The final information is displayed through the Streamlit dashboard and can be exported as a PDF report.

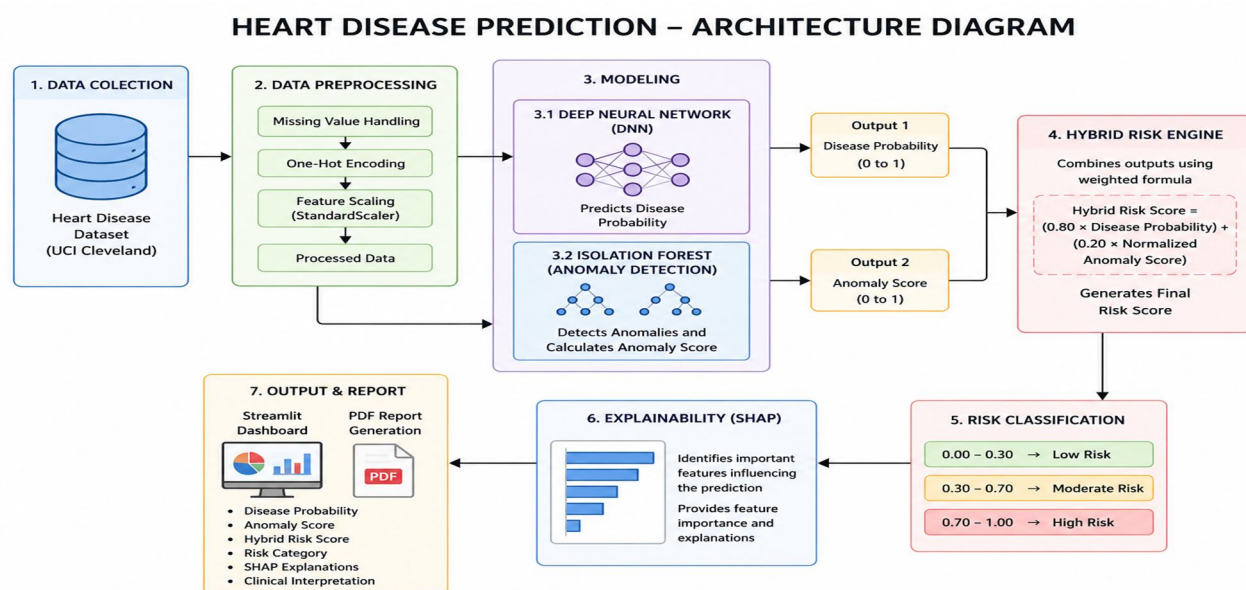


Fig. 1. Final System Architecture of the Proposed Hybrid Cardiovascular Disease Prediction Framework

VI. IMPLEMENTATION

The system was implemented using Python-based machine learning and deep learning technologies. TensorFlow/Keras is used for the trained DNN, Scikit-learn is used for preprocessing and Isolation Forest, Pandas and NumPy support data processing, SHAP provides explainability, Streamlit provides the interactive dashboard, Joblib stores and loads preprocessing and anomaly-detection objects, and ReportLab supports automated PDF generation.

Technology	Purpose
Python	System development
TensorFlow/Keras	Deep Neural Network
Scikit-learn	Isolation Forest and preprocessing
Pandas	Data processing
NumPy	Numerical computation
SHAP	Explainable AI
Streamlit	Interactive dashboard
Joblib	Model/scaler serialization
ReportLab	PDF report generation

Table III. Technologies Used in the Implementation

VII. RESULTS AND DISCUSSION

A. Patient-Level Prediction Example

During system testing, one example patient input produced a disease probability of 72.95%, a raw anomaly-related score of 0.0558, and a hybrid risk score of 68.08%. According to the implemented thresholds, the corresponding hybrid category is Moderate Risk. These values demonstrate the output generated by the deployed application for an individual input.

Metric	Example Value
Disease Probability	72.95%
Anomaly-Related Score	0.0558
Hybrid Risk Score	68.08%
Risk Category	Moderate Risk

Table IV. Example Patient-Level Prediction Output

The example result should not be interpreted as overall model accuracy or population-level validation performance. The disease probability is generated by the trained DNN, while the anomaly component provides supplementary information about the atypicality of the input pattern. The hybrid score therefore represents an application-level combination of predictive probability and anomaly awareness.

B. Model Evaluation Reporting

The final evaluation artifacts provide aggregate performance results for the held-out test set. From the confusion matrix (TN = 67, FP = 15, FN = 12, TP = 90), the DNN achieved 85.33% accuracy, 85.71% precision, 88.24% recall, and an F1-score of 86.96%. The ROC curve reports an ROC-AUC of 0.913. These values represent model-level test performance and are distinct from the patient-level example shown above.

Metric	Final Value
Test-set size	184
True Negatives / False Positives	67 / 15
False Negatives / True Positives	12 / 90
Accuracy	85.33%
Precision	85.71%
Recall	88.24%
F1-score	86.96%
ROC-AUC	0.913

Table V. Final Measured Model Performance

The implemented application produces an individual cardiovascular risk assessment consisting of disease probability, anomaly score, hybrid risk score, and risk category. One example generated during system testing produced a disease probability of 72.95%, an anomaly score of 0.0558, and a hybrid risk score of 68.08%. According to the implemented risk thresholds, the corresponding category is Moderate Risk.

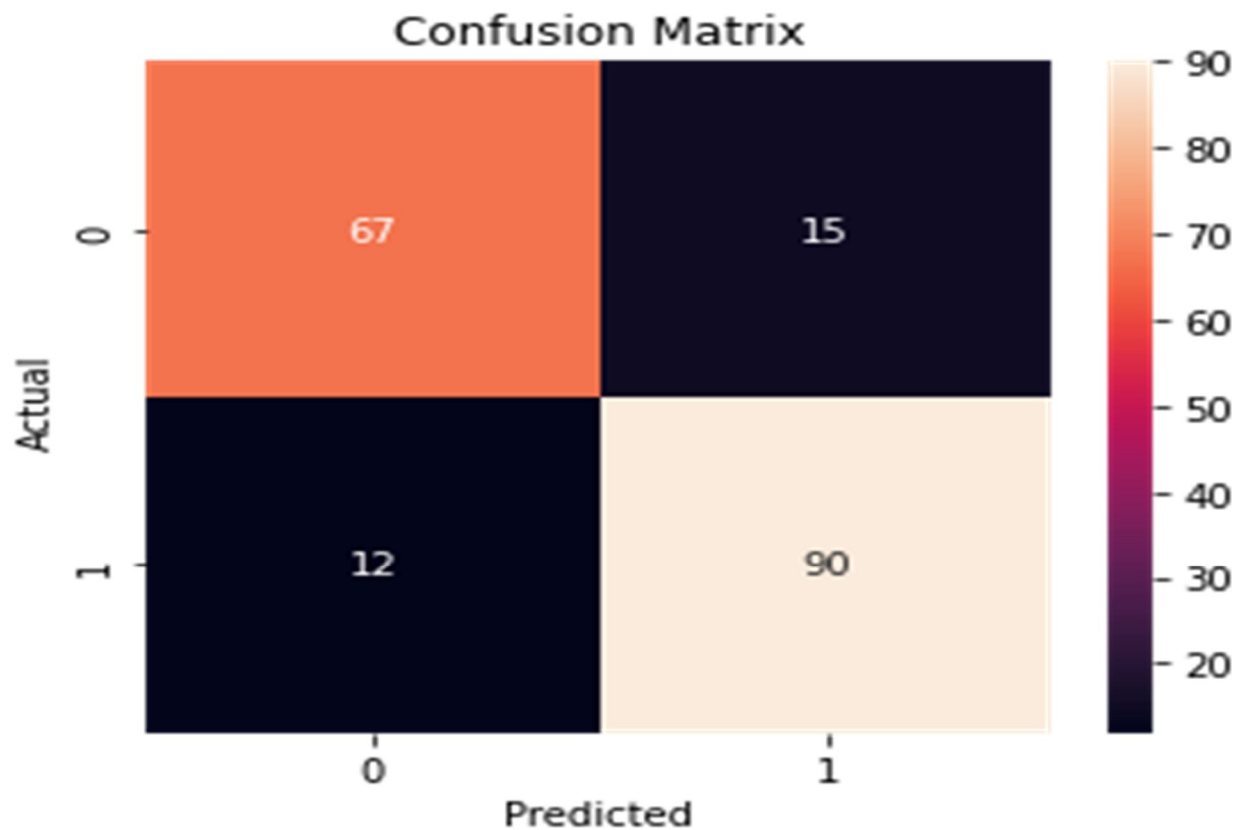


Fig. 2. Confusion Matrix for Final Test-Set Evaluation

These values represent an example individual prediction and should not be interpreted as overall model accuracy or validation performance. Overall performance measures such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix should be added to the paper only after they are obtained from the actual trained-model evaluation.

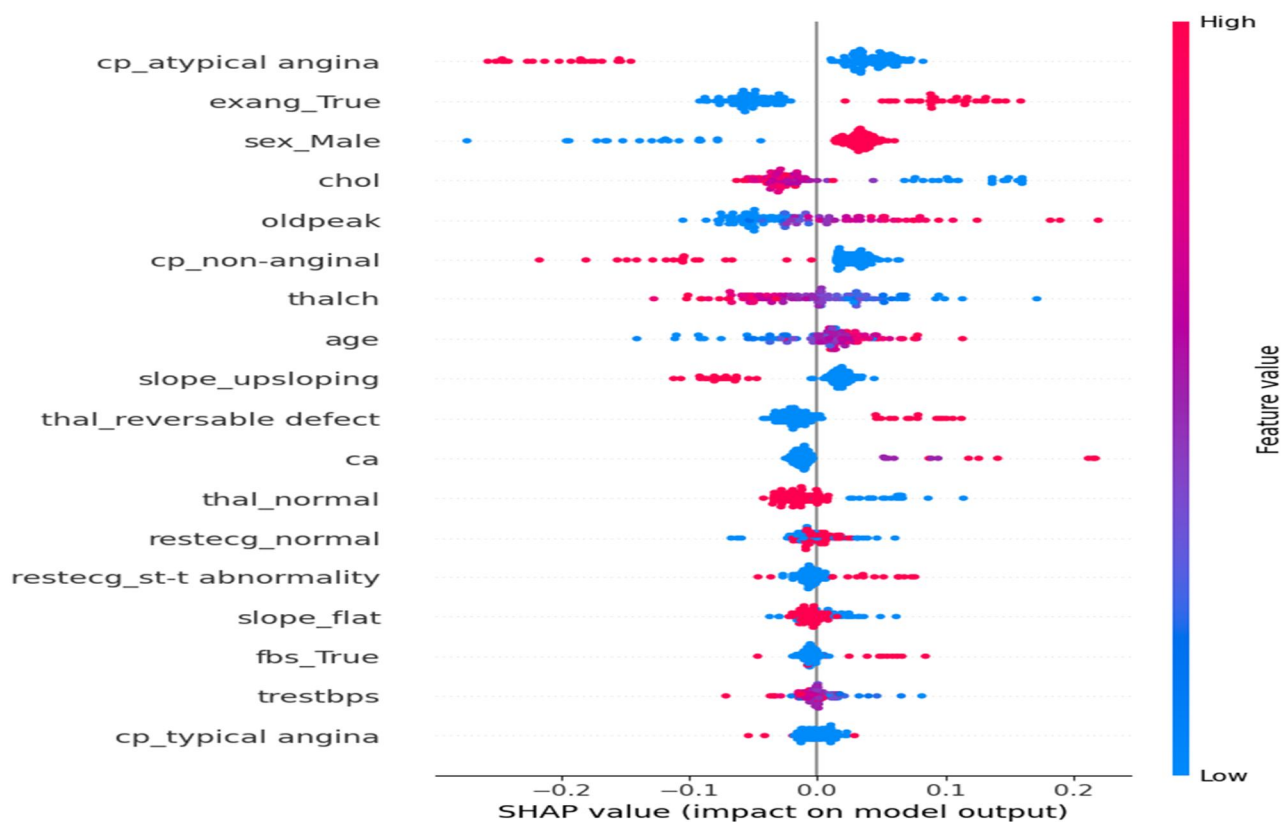


Fig. 3. ROC Curve for the Final DNN Evaluation (ROC-AUC = 0.913)

The confusion matrix indicates 157 correct predictions out of 184 test samples. The relatively high recall (88.24%) shows that the model identified a large proportion of positive cases, while the precision (85.71%) indicates that a strong proportion of predicted positive cases were correct. The ROC-AUC of 0.913 indicates good discrimination between the two classes on the evaluated test set.

Example Prediction Output

Metric	Example Value
Disease Probability	72.95%
Anomaly Score	0.0558
Hybrid Risk Score	68.08%
Risk Category	Moderate Risk

VIII. EXPLAINABILITY ANALYSIS

The SHAP component provides an explanation of the model's individual prediction. The generated PDF reports show the most influential features together with their SHAP values and whether their contribution is interpreted as increasing or reducing the model output. This enables users to move beyond a single probability value and inspect the clinical variables that influenced the prediction.

For example, the implemented reporting workflow has displayed features including Thalassemia (Normal), Gender (Male), ST Segment Slope, Exercise-Induced Angina, and Number of Major Vessels. The exact ranking varies with the patient input and model explanation.

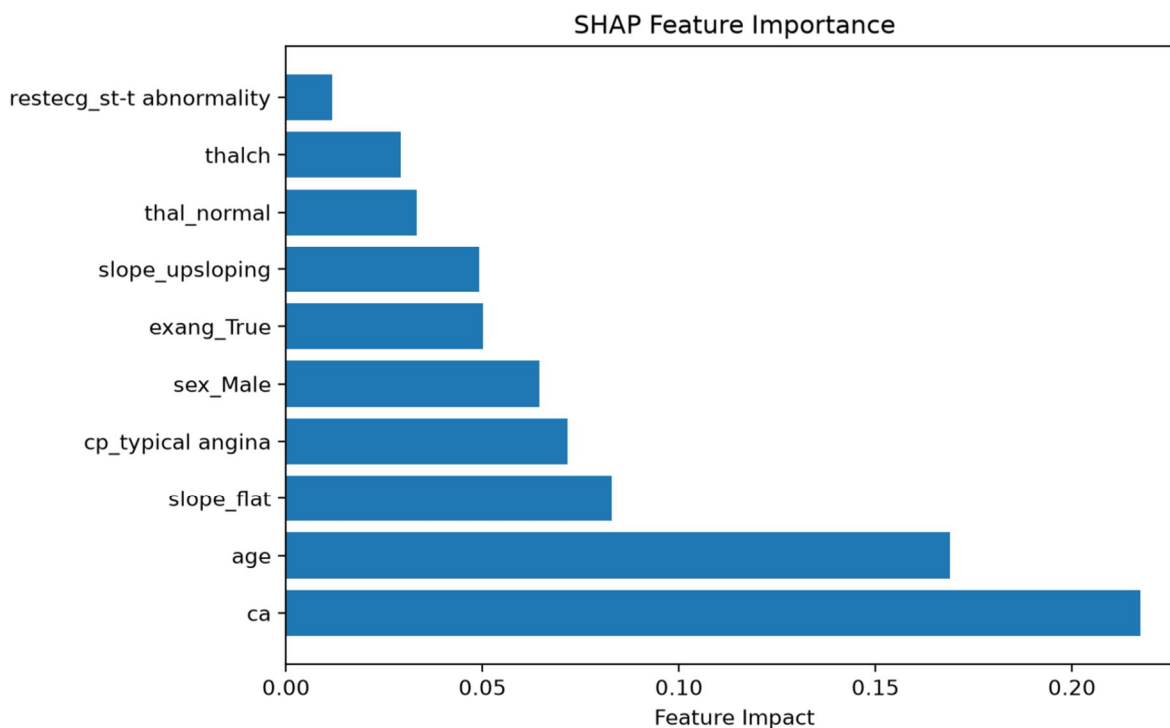


Fig. 2. SHAP Feature Importance for the Heart Disease Prediction Model

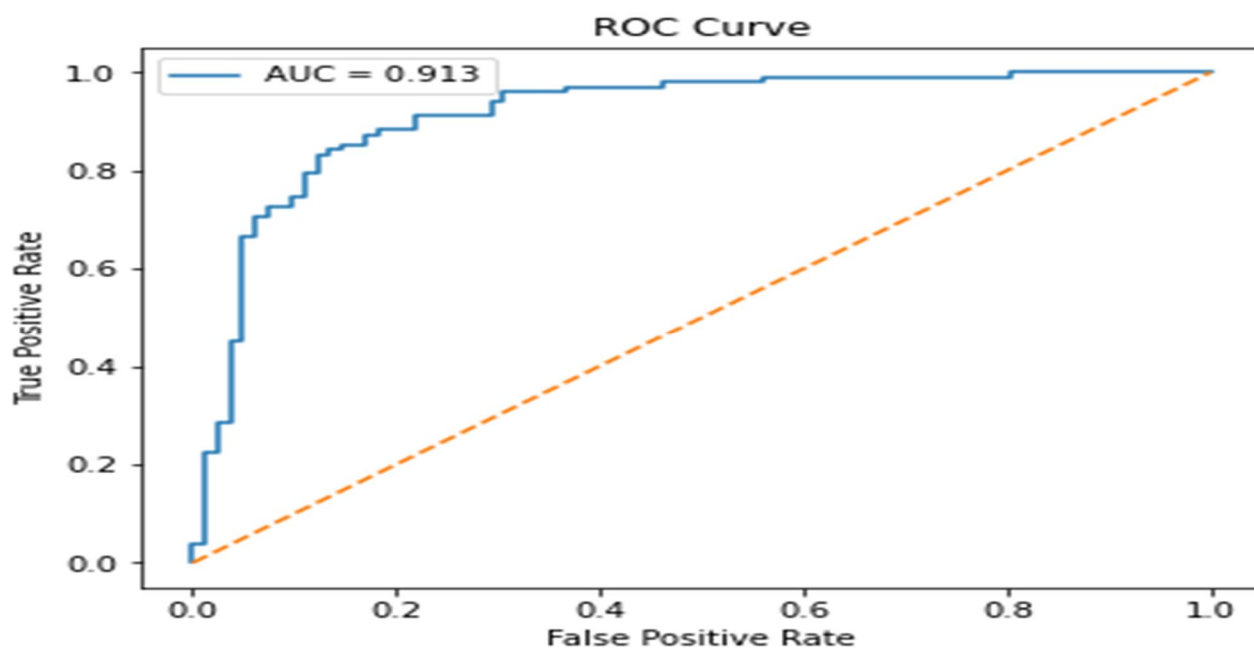


Fig. 5. SHAP Summary Plot Showing Feature-Level Contributions

The global SHAP feature-importance plot summarizes the overall magnitude of influence of the encoded input features. The SHAP summary plot additionally shows how low and high feature values contribute positively or negatively to the model output across the evaluated samples. Together, these visualizations complement the patient-specific SHAP values displayed in the dashboard and PDF report.

IX. USER INTERFACE AND AUTOMATED REPORTING

The Streamlit application provides an interactive interface for entering patient attributes and obtaining a prediction. The dashboard displays the disease probability, anomaly score, hybrid risk score, risk category, and SHAP feature importance. A PDF report is generated to summarize the prediction, risk category, feature-level explanations, and clinical interpretation.

X. LIMITATIONS

The current implementation has several limitations. First, it is developed using the UCI Cleveland Heart Disease Dataset, which is a structured clinical dataset and may not represent the full diversity of real-world patient populations. Second, the system uses a fixed set of clinical variables and does not incorporate ECG signals, medical images, longitudinal electronic health records, or wearable-device data. Third, the hybrid-score weights are fixed at 80% for DNN probability and 20% for the normalized anomaly component. Finally, the system is a research and decision-support prototype and should not be treated as a standalone clinical diagnostic tool.

XI. FUTURE SCOPE

Future development can focus on larger and more diverse cardiovascular datasets, external validation using independent datasets, optimization of hybrid-score weights, integration of ECG and medical imaging data, real-time wearable-device data, longitudinal patient monitoring, cloud deployment, advanced deep-learning architectures, and more extensive clinical validation.

XII. CONCLUSION

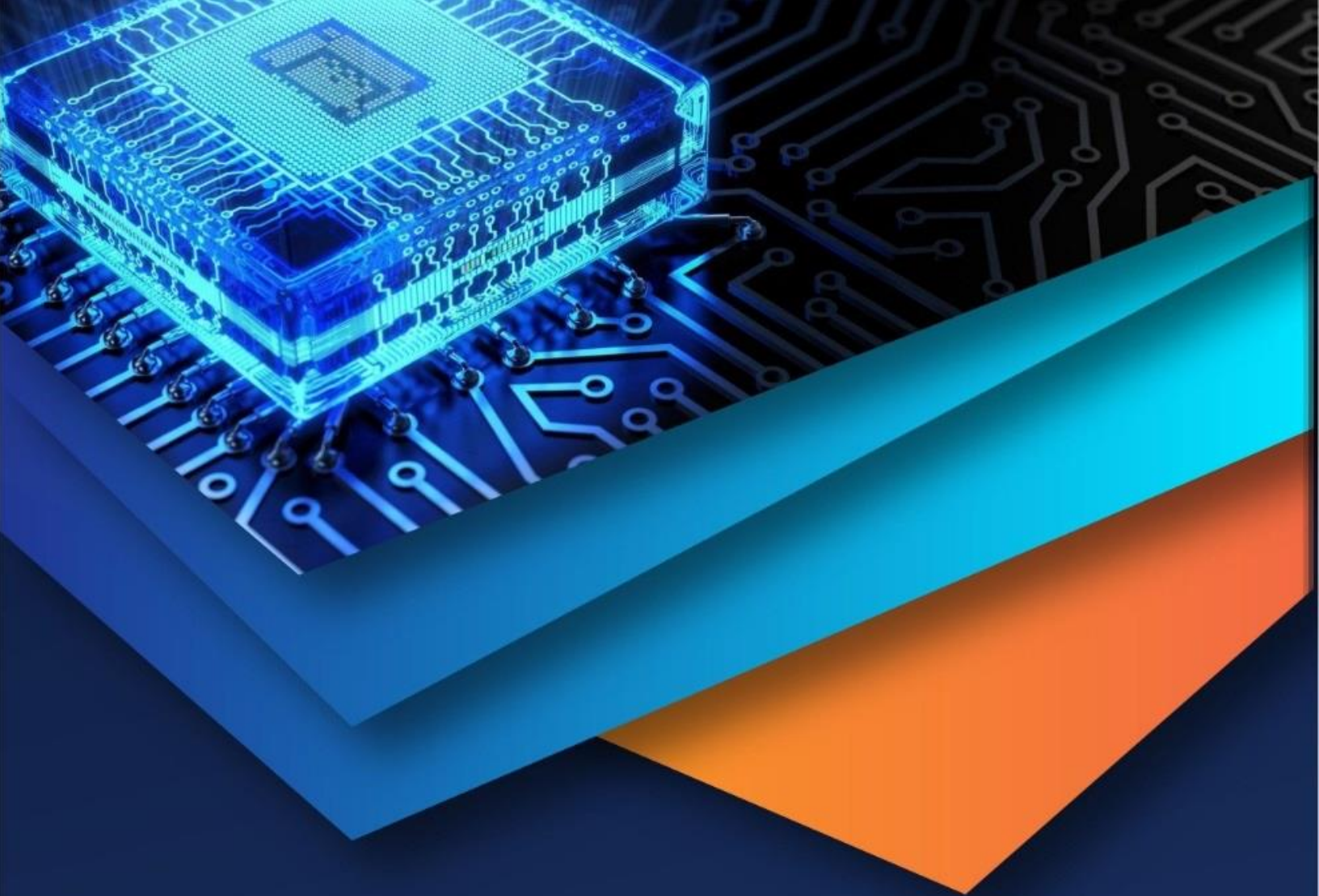
This paper presents an adaptive hybrid cardiovascular disease prediction framework that integrates Deep Neural Networks, Isolation Forest anomaly detection, hybrid risk scoring, and SHAP Explainable AI. The DNN provides disease probability estimation, while the Isolation Forest adds anomaly awareness to the prediction workflow. The outputs are combined through a hybrid risk score and categorized into low, moderate, and high-risk levels.

The integration of SHAP provides feature-level explanations, improving the interpretability of the prediction. Furthermore, the Streamlit dashboard and automated PDF reporting module transform the machine-learning pipeline into an interactive application capable of presenting prediction results and explanations in a user-friendly format.

The implemented framework therefore provides an integrated approach combining prediction, anomaly detection, explainability, and reporting for cardiovascular disease risk assessment.

REFERENCES

- [1] Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H., Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J. L., & Tang, W. H. W. (2020). Machine learning prediction in cardiovascular diseases: A meta-analysis. *Scientific Reports*, 10(1), Article 16057.
- [2] Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access*, 9, 106575–106588.
- [3] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. M. J. M., Ignatious, E., Shultana, S., Beeravolu, A. R., & de Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques. *IEEE Access*, 9, 19304–19326.
- [4] Li, Q., Campan, A., Ren, A., & Eid, W. E. (2022). Automating and improving cardiovascular disease prediction using machine learning and EMR data features from a regional healthcare system. *International Journal of Medical Informatics*, 163, Article 104786.
- [5] Rahmathullah, M., Bhavanam, S. N., & Midasala, V. (2025). Deep learning-powered cardiovascular disease prediction: A novel approach to early diagnosis and risk assessment. *Journal of Neonatal Surgery*, 14(4), 21–31.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)