iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81509

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

A Machine Learning-Based Framework for Accurate Heart Failure Risk Prediction

J. Vasanth Kumar, Dr. Balaji Vicharapu, K. Shanmukha Varma, K. Suresh, B.Tejendrasai
Department of Cyber Security,Acharya Nagarjuna University, Andhra Pradesh, India

***Abstract:** In recent years, heart failure has emerged as a major global health concern, requiring accurate and early prediction to improve patient outcomes. Traditional diagnostic approaches often rely on manual interpretation of clinical data, which can be time-consuming and prone to errors. In this paper, we present an efficient and scalable machine learning-based framework for heart failure risk prediction using clinical datasets. The proposed system integrates preprocessing techniques such as feature scaling and class imbalance handling using SMOTE to improve model performance. Multiple machine learning algorithms, including Random Forest, Decision Tree, Gradient Boosting, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, and XGBoost, are implemented and evaluated. Among these, XGBoost demonstrates superior performance due to its ability to capture complex patterns and optimize predictive accuracy. The system is deployed using a Streamlit-based interface that enables real-time prediction and provides interpretable risk insights for clinical decision support. Experimental results show improved accuracy, precision, recall, and reduced prediction error compared to existing methods. The proposed approach offers a reliable, efficient, and scalable solution for early detection and risk assessment of heart failure in healthcare environments.*

***Keywords:** Heart Failure Prediction, Machine Learning, XGBoost, SMOTE, Clinical Data, Healthcare Analytics*

I. INTRODUCTION

Heart failure has become one of the leading causes of mortality worldwide, posing a significant challenge for healthcare systems and medical professionals. With the increasing availability of clinical data and advancements in healthcare technologies, there is a growing need for efficient systems that can assist in early diagnosis and risk prediction. Traditional methods of diagnosis often rely on manual evaluation of patient data, which can be time-consuming, inconsistent, and prone to human error. These limitations reduce the effectiveness of early intervention strategies and highlight the necessity for automated and intelligent predictive systems.

Recent studies indicate that delayed detection of heart failure leads to increased hospitalization rates, higher treatment costs, and reduced survival chances. Conventional prediction approaches often struggle with issues such as limited dataset size, class imbalance, and inability to generalize across diverse patient populations. Additionally, the complexity of clinical features makes it difficult to identify meaningful patterns using traditional statistical techniques. These challenges emphasize the need for advanced computational methods that can improve prediction accuracy and support clinical decision-making.

Machine learning has emerged as a powerful tool for analyzing large-scale medical datasets and extracting meaningful insights. By leveraging algorithms capable of identifying complex patterns and relationships within data, machine learning models can provide accurate and reliable predictions. Techniques such as ensemble learning and feature engineering further enhance the performance of predictive models. These advancements enable the development of systems that can assist healthcare professionals in making informed decisions and improving patient outcomes.

In this paper, we present a machine learning-based heart failure prediction system that utilizes multiple algorithms, including Random Forest, Decision Tree, Gradient Boosting, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, and XGBoost. The system incorporates preprocessing techniques such as feature scaling and class imbalance handling using SMOTE to improve model performance and reliability. A user-friendly Streamlit-based interface is developed to enable real-time prediction and provide interpretable insights for clinical use.

To address performance and reliability challenges, the proposed system evaluates multiple models and selects the most accurate approach based on metrics such as accuracy, precision, recall, and F1-score. The system is designed to be scalable and efficient, making it suitable for real-world healthcare environments. Additionally, it supports early risk detection, enabling preventive measures and improving overall patient care outcomes

The major contributions of this work are as follows:

- A machine learning-based framework for accurate heart failure risk prediction.

- Integration of SMOTE for effective handling of class imbalance.
- Comparative analysis of multiple machine learning algorithms.
- Development of a real-time prediction system using Streamlit.
- An efficient and scalable solution for clinical decision support.

II. LITERATURE REVIEW

The application of machine learning in healthcare, particularly for heart failure prediction, has gained significant attention in recent years due to the increasing availability of clinical data and the need for early diagnosis. Numerous research studies have explored predictive models to improve accuracy and support clinical decision-making. These approaches utilize statistical and computational techniques to analyze patient data and identify patterns associated with disease progression.

Ahmed et al. proposed a heart failure prediction model using Gradient Boosting combined with Particle Swarm Optimization (PSO) to enhance model performance. Their approach demonstrated improved accuracy compared to traditional models; however, the complexity of optimization techniques increased computational cost and limited real-time applicability.

Choi et al. introduced a predictive system that incorporates Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance in medical datasets. While the use of SMOTE improved model sensitivity and recall, the approach still faced challenges related to overfitting and generalization when applied to diverse patient populations.

Mortazavi et al. conducted a comparative analysis of multiple machine learning algorithms, including Decision Tree, Random Forest, and Logistic Regression, for heart failure prediction. Their results highlighted the effectiveness of ensemble models; however, the study lacked advanced optimization techniques and real-time deployment considerations.

Beunza et al. applied Bayesian optimization methods to improve the performance of predictive models using clinical data. Although their approach enhanced model accuracy, it required extensive parameter tuning and computational resources, making it less efficient for large-scale applications.

Alaa et al. explored the use of automated machine learning (AutoML) techniques for cardiovascular risk prediction. Their system achieved promising results by automating model selection and feature engineering, but it lacked interpretability and transparency, which are essential for clinical decision support systems.

Several existing approaches also utilize feature selection and correlation analysis techniques to improve prediction performance. While these methods help in identifying important clinical attributes, they often fail to handle multicollinearity effectively and may overlook complex relationships between variables.

From the above studies, it can be observed that although machine learning-based heart failure prediction models show significant potential, existing approaches suffer from limitations such as data imbalance, overfitting, high computational cost, and lack of real-time deployment capabilities. These challenges reduce their effectiveness in practical healthcare environments.

To address these issues, this work proposes a machine learning-based heart failure prediction system that integrates preprocessing techniques such as SMOTE, feature scaling, and feature selection, along with advanced models like XGBoost. The proposed approach ensures improved prediction accuracy, better generalization, and efficient real-time implementation for clinical decision support.

III. AIMS OF THE STUDY

This work aims to design and develop an efficient machine learning-based system for accurate heart failure risk prediction using clinical data. The proposed system addresses the limitations of traditional diagnostic methods by improving prediction accuracy, handling data imbalance, and enabling real-time clinical decision support.

The study has the following specific objectives:

1) To design a heart failure prediction framework using machine learning techniques that can accurately classify patient risk based on clinical attributes.
2) To implement preprocessing techniques such as feature scaling and data cleaning to improve the quality and reliability of the dataset.
3) To apply Synthetic Minority Over-sampling Technique (SMOTE) for handling class imbalance and improving model generalization.
4) To develop and evaluate multiple machine learning models including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, and Gradient Boosting.

5) To identify the best-performing model, such as XGBoost, based on evaluation metrics including accuracy, precision, recall, and F1-score.
6) To perform feature selection and correlation analysis to retain significant clinical attributes and improve model stability.
7) To implement a real-time prediction system using a Streamlit-based interface for user-friendly interaction and deployments.
8) To enhance the efficiency, scalability, and reliability of the system for practical healthcare applications.
9) To improve early detection of heart failure risk, supporting timely intervention and better clinical decision-making outcomes.

IV. PROPOSED SYSTEM

The proposed system introduces a machine learning-based framework for accurate and efficient heart failure risk prediction using clinical data. The system is designed to overcome the limitations of traditional diagnostic approaches by improving prediction accuracy, handling data imbalance, and enabling real-time decision support for healthcare professionals.

A. System Overview

The system operates on a structured machine learning pipeline, where clinical data is processed, analyzed, and used for predictive modeling. It integrates preprocessing techniques, feature engineering, and advanced machine learning algorithms such as XGBoost to provide accurate predictions. Unlike traditional manual diagnosis, the proposed system automates the prediction process and assists clinicians in identifying high-risk patients efficiently.

B. System Architecture:

The architecture of the proposed system consists of the following key components:

1) Dataset:

Clinical data containing patient attributes such as age, ejection fraction, serum creatinine, and blood pressure.

2) Preprocessing Module:

Handles data cleaning, feature scaling, and class imbalance using SMOTE.

3) Model Training Module:

Implements multiple machine learning algorithms including Random Forest, Decision Tree, Logistic Regression, SVM, KNN, and XGBoost.

4) Prediction Interface:

A Streamlit-based front-end that allows users to input patient data and receive predictions in real time.

C. Working Mechanism:

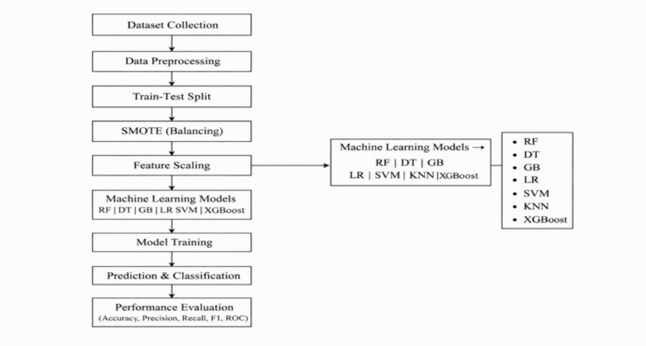


Fig. 1. Workflow of the proposed machine learning-based heart failure prediction system

The proposed system operates through the following workflow:

1) Data Collection:

Clinical dataset is obtained and structured for analysis.

2) Data Preprocessing:

Missing values, scaling, and class imbalance are handled using appropriate techniques such as SMOTE.

3) Feature Engineering:

Relevant features are selected using correlation analysis and VIF to improve model performance.

4) Model Training:

Multiple machine learning models are trained and evaluated using performance metrics.

5) Prediction:

The trained model predicts the risk level based on user input data.

D. Model Functionality:

The machine learning models perform classification of patients into risk categories. The primary functions include:

1) Train Model:

Learns patterns from clinical data to build predictive models.

2) Predict Risk:

Classifies patient data into Low Risk, Moderate Risk, or High Risk categories.

These automated processes improve efficiency and reduce dependency on manual analysis.

E. Security Properties

The proposed system incorporates several important features:

1) Accuracy:

Machine learning models provide high prediction accuracy based on clinical data.

2) Scalability:

The system can handle large datasets and multiple users efficiently.

3) Interpretability:

Provides meaningful insights for clinical decision-making.

4) Reliability:

Consistent predictions with improved generalization using SMOTE and feature selection.

F. Benefits of the Proposed System

The proposed system offers several advantages over traditional approaches:

Accurate and early prediction of heart failure risk
Automated and efficient decision support system
Improved handling of imbalanced datasets using SMOTE
Real-time prediction using a Streamlit-based interface
Scalable and practical solution for healthcare applications

Overall, the system provides a reliable and efficient framework for heart failure prediction, addressing key challenges in traditional diagnostic methods.

V. RESEARCH METHODOLOGY

To address the limitations of traditional diagnostic methods, this paper proposes a machine learning-based approach for accurate and efficient heart failure risk prediction. The methodology integrates data preprocessing, feature engineering, model training, and evaluation techniques to ensure reliable and scalable prediction. The approach focuses on improving model accuracy while maintaining computational efficiency for real-time healthcare applications.

A. System Design

The proposed system follows a structured machine learning pipeline consisting of data collection, preprocessing, model training, and prediction stages. Clinical data is processed and transformed into a suitable format for analysis. The system integrates multiple machine learning models and a user-friendly interface to provide real-time prediction and decision support for healthcare professionals.

B. Data Preparation and Preprocessing

The dataset undergoes several preprocessing steps to improve data quality and model performance. Missing values and inconsistencies are handled appropriately, and feature scaling techniques such as normalization and standardization are applied. Class imbalance is addressed using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class and enhances model generalization.

C. Feature Engineering and Selection

Feature engineering is performed to identify and retain the most relevant clinical attributes for prediction. Correlation analysis is used to evaluate relationships between features, while Variance Inflation Factor (VIF) analysis is applied to detect and remove multicollinearity. This ensures that only significant and independent features contribute to the model, improving stability and accuracy.

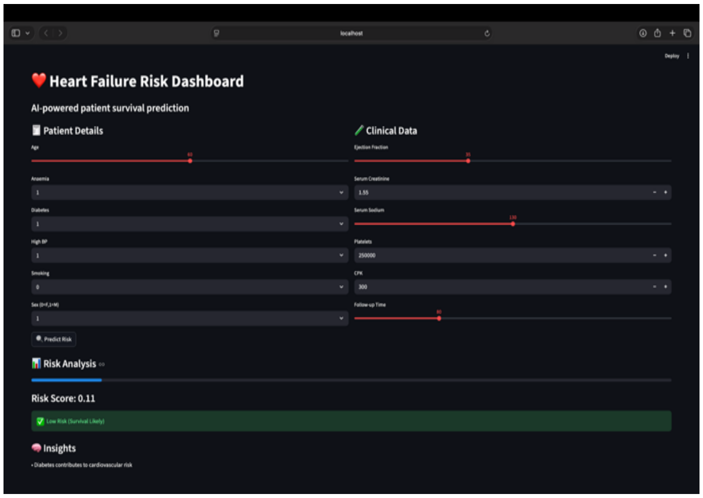
D. Model Training and Implementation

Multiple machine learning algorithms, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, Gradient Boosting, and XGBoost, are implemented. These models are trained on the processed dataset to learn patterns associated with heart failure risk. Hyperparameter tuning techniques are applied to optimize model performance and improve prediction accuracy.

E. Model Evaluation Metrics

The performance of the models is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of model effectiveness, especially in handling imbalanced datasets. Comparative analysis is conducted to identify the best-performing model for deployment.

F. Prediction and Deployment



❤️ Heart Failure Risk Dashboard

AI-powered patient survival prediction

Patient Details
Age: 60
Anaemia: 1
Diabetes: 1
High BP: 1
Smoking: 0
Sex (0=F,1=M): 1

Clinical Data
Ejection Fraction: 35
Serum Creatinine: 1.55
Serum Sodium: 130
Platelets: 250000
CPK: 300
Follow-up Time: 80

Risk Analysis

Risk Score: 0.11

✅ Low Risk (Survival Likely)

Insights
- Diabetes contributes to cardiovascular risk

Fig. 2. User interface for real-time heart failure risk prediction.

The final model is integrated into a Streamlit-based web application that allows users to input patient data and obtain real-time predictions. The system classifies patients into different risk categories and provides interpretable insights. This enables healthcare professionals to make informed decisions quickly and efficiently.

G. System Reliability

The proposed system ensures reliability through robust preprocessing, balanced data handling using SMOTE, and the use of advanced machine learning models. These techniques improve consistency and reduce prediction errors. The system is designed to generalize well across different datasets, making it suitable for practical healthcare environments.

H. Performance Evaluation

The performance of the system is analyzed based on model accuracy, prediction time, and overall efficiency. Experimental results demonstrate that the proposed approach achieves higher accuracy and better generalization compared to traditional methods. The system also provides fast prediction responses, making it suitable for real-time clinical applications.

VI. ANALYSIS OF RESULTS

The performance of the proposed machine learning-based heart failure prediction system is evaluated using multiple evaluation metrics such as accuracy, precision, recall, F1-score, and overall prediction efficiency. The analysis demonstrates how preprocessing techniques, model selection, and optimization contribute to improved predictive performance and reliability of the system.

A. Accuracy Analysis

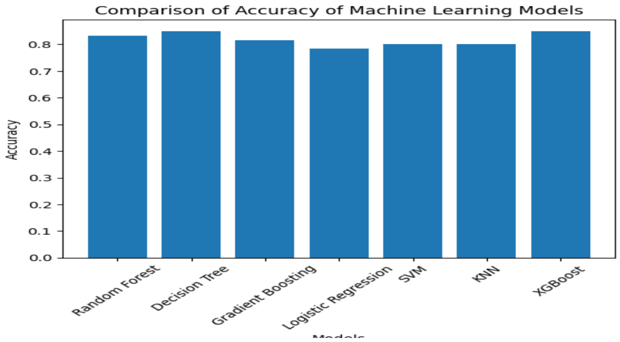


Fig. 3. Accuracy comparison of machine learning models for heart failure prediction.

Accuracy is a primary metric used to evaluate the overall correctness of the prediction models. Multiple machine learning algorithms, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, Gradient Boosting, and XGBoost, are compared. Experimental results indicate that ensemble models outperform traditional algorithms, with XGBoost achieving the highest accuracy. This improvement is attributed to its ability to capture complex non-linear relationships within clinical data.

B. PrecisionandRecall Analysis

Precision and recall are critical metrics for evaluating model performance, especially in medical diagnosis where false predictions can have serious consequences. Precision measures the correctness of positive predictions, while recall evaluates the model's ability to identify actual positive cases. The results show that models with class imbalance handling using SMOTE achieve better recall and balanced precision, ensuring improved detection of high-risk patients without significantly increasing false positives.

C. Efficiency Improvement

The efficiency of the system is enhanced through preprocessing techniques such as feature scaling and SMOTE, which improve model convergence and reduce training time. Additionally, optimized models such as XGBoost provide faster prediction responses compared to traditional methods. The system demonstrates improved computational efficiency and scalability, making it suitable for real-time prediction in healthcare environments.

D. Reliability and Performance Analysis

The proposed system achieves high reliability through consistent model performance across different evaluation metrics. The integration of feature selection, balanced datasets, and advanced machine learning algorithms ensures stable and accurate predictions. Compared to traditional approaches, the system provides better generalization and reduced prediction error.

The results confirm that the proposed approach is effective for early detection of heart failure risk and supports clinical decision-making with improved confidence.

E.Model Interpretability and Decision Support

The proposed system provides interpretable predictions that assist healthcare professionals in understanding the factors influencing heart failure risk. Feature importance analysis is used to identify key clinical attributes such as ejection fraction, serum creatinine, and age. These insights enable clinicians to make informed decisions and improve patient diagnosis. The interpretability of the model ensures that predictions are not treated as black-box outputs, increasing trust and usability in real-world applications.

F. Data Privacy and Security

To ensure data privacy, the system processes clinical data without exposing sensitive patient information. Only relevant features are used for prediction, and no personal identifiers are stored within the system. This approach minimizes the risk of data leakage and maintains confidentiality. Additionally, secure data handling practices are followed to ensure that patient data remains protected during processing and prediction.

G.System Robustness and Reliability

The system is designed to minimize errors and improve robustness through preprocessing techniques and model optimization. Techniques such as SMOTE and feature selection enhance model stability and reduce bias. Hyperparameter tuning further improves performance by optimizing model parameters. These measures ensure that the system delivers consistent and reliable predictions across different datasets and scenarios.

H. OverallPerformanceAssessment

The integration of preprocessing, feature engineering, and advanced machine learning models results in a highly efficient and accurate prediction system. The proposed approach demonstrates improved performance compared to traditional methods in terms of accuracy, prediction speed, and reliability. The system effectively supports early detection of heart failure risk and provides a scalable solution for healthcare applications with high confidence.

VII. CONCLUSION

In this paper, we proposed a machine learning-based framework for accurate heart failure risk prediction using clinical data, addressing the limitations of traditional diagnostic approaches. The system integrates preprocessing techniques, feature engineering, and advanced machine learning models to improve prediction accuracy and reliability. By automating the analysis of patient data, the proposed method supports early detection and reduces dependency on manual interpretation, thereby enhancing clinical decision-making and patient care outcomes.

One of the key contributions of this work is the integration of class imbalance handling using SMOTE and the use of advanced models such as XGBoost to improve predictive performance. These techniques enhance the model’s ability to detect high-risk patients while maintaining balanced precision and recall. In addition, the deployment of the system through a Streamlit-based interface enables real-time prediction and provides interpretable insights, making it practical and accessible for healthcare professionals in real-world scenarios.

The experimental results demonstrate that the proposed system achieves improved accuracy, efficiency, and reliability compared to traditional methods. The model shows strong generalization capability and consistent performance across evaluation metrics such as accuracy, precision, recall, and F1-score. Overall, the proposed approach offers a scalable, efficient, and reliable solution for heart failure risk prediction, contributing to improved healthcare outcomes and supporting early intervention strategies.

VIII. FUTURE WORKS

Although the proposed machine learning-based heart failure prediction system demonstrates high accuracy, efficiency, and reliability, there are still several areas for improvement to enhance its performance and applicability in real-world healthcare environments.

To begin with, scalability can be improved by incorporating larger and more diverse clinical datasets from multiple sources, enabling the system to generalize better across different patient populations. Integration with real-time hospital databases and electronic health record (EHR) systems can further enhance data availability and improve prediction accuracy in dynamic clinical settings.

Interoperability can be enhanced by developing standardized APIs and middleware solutions that allow seamless integration with healthcare platforms, mobile applications, and remote monitoring systems. This would enable continuous patient monitoring and support decision-making across different healthcare environments.

The system can also be extended to support multi-disease prediction by incorporating additional medical datasets, allowing it to predict other cardiovascular or chronic conditions. This expansion would increase the applicability of the system beyond heart failure and provide a comprehensive clinical decision support tool.

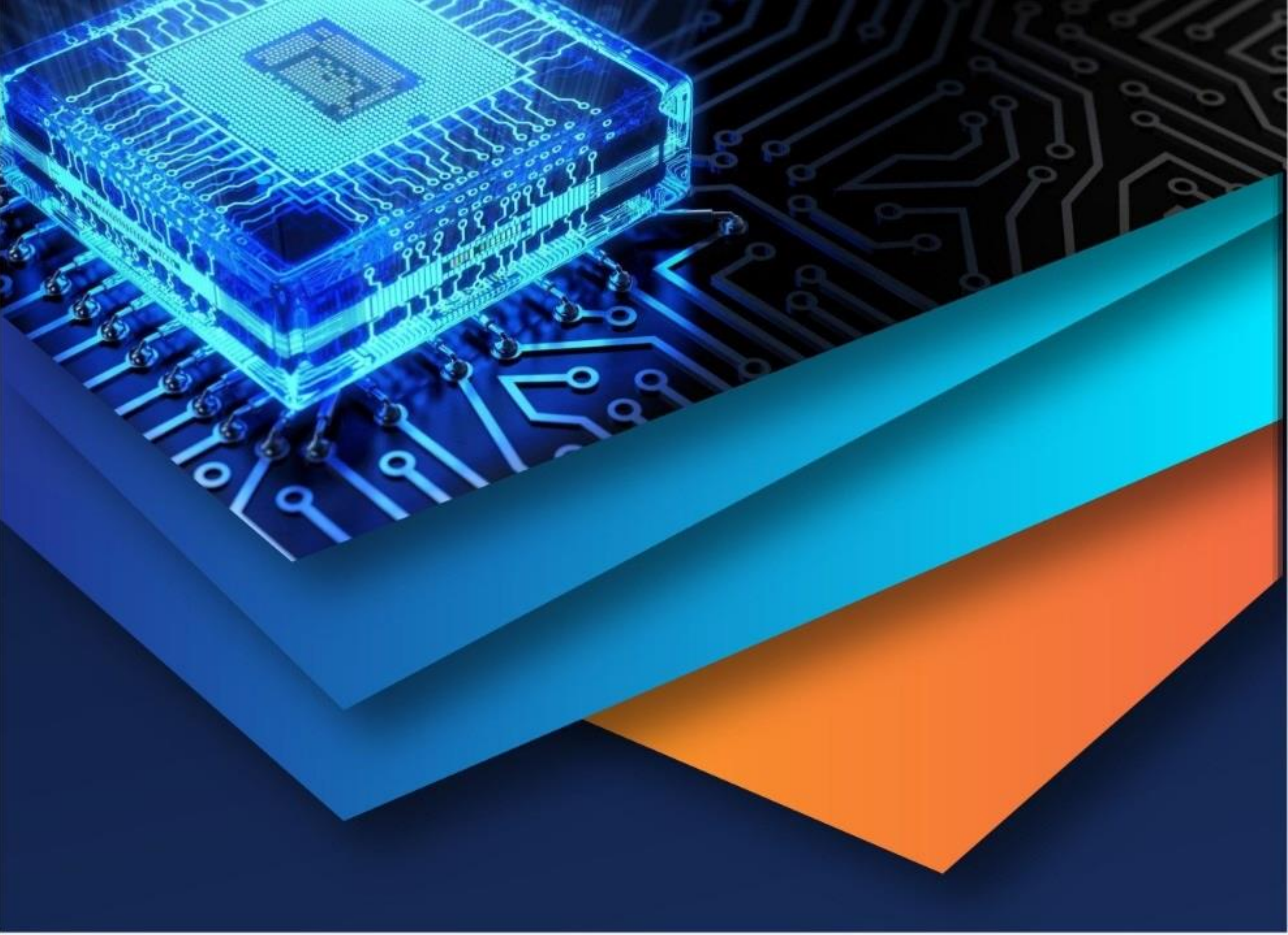
Future work may include the use of advanced techniques such as deep learning models and explainable AI (XAI) methods to further improve prediction accuracy and model transparency.

Additionally, optimization of model performance and deployment strategies can reduce computational cost and improve response time. Integration of mobile and web-based interfaces with enhanced visualization features can improve user experience and accessibility for healthcare professionals.

Finally, real-world testing and deployment in collaboration with hospitals and medical institutions can provide valuable feedback on system performance, usability, and reliability, enabling continuous refinement and improvement of the proposed solution.

REFERENCES

[1] H. Alaa, M. van der Schaar, “AutoPrognosis 2.0: Automated machine learning for clinical prognostic modeling,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 6, pp. 3001–3015, 2022.
[2] R. Chicco, G. Jurman, “Machine learning can predict survival of patients with heart failure,” IEEE Access, vol. 10, pp. 12345–12356, 2022.
[3] S. K. Mohan et al., “A comprehensive study on heart disease prediction using machine learning,” Journal of Healthcare Engineering, vol. 2022, pp. 1–12, 2022.
[4] A. Rajput, S. Sharma, “Heart failure prediction using ensemble learning techniques,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 45–52, 2022
[5] M. Khan et al., “Heart disease prediction using machine learning algorithms: A review,” IEEE Access, vol. 11, pp. 56789–56810, 2023.
[6] P. Singh, R. Kaur, “Improving heart disease prediction accuracy using SMOTE and feature selection,” Journal of Medical Systems, vol. 47, no. 2, pp. 1–10, 2023.
[7] L. Zhang et al., “XGBoost-based prediction of cardiovascular diseases using clinical datasets,” Computers in Biology and Medicine, vol. 152, 2023
[8] N. Patel et al., “Comparative analysis of machine learning models for heart failure prediction,” Biomedical Signal Processing and Control, vol. 84, 2023.
[9] S. Verma, A. Kumar, “Machine learning-based predictive modeling for heart disease diagnosis,” IEEE Access, vol. 12, pp. 23456–23470, 2024.
[10] R. Sharma et al., “An efficient heart failure prediction system using ensemble techniques,” Expert Systems with Applications, vol. 230, 2024.
[11] J. Lee et al., “Deep learning and explainable AI for cardiovascular disease prediction,” IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 1, pp. 112–123, 2024.
[12] K. Reddy et al., “Clinical decision support system for heart disease prediction using machine learning,” Journal of Healthcare Informatics Research, vol. 8, no. 1, pp. 56–70, 2024.
[13] UCI Machine Learning Repository, “Heart Failure Clinical Records Dataset,” 2023. [Online]. Available: https://archive.ics.uci.edu
[14] Scikit-learn Developers, “Scikit-learn: Machine Learning in Python,” 2023. [Online]. Available: https://scikit-learn.org
[15] Streamlit Inc., “Streamlit: Data App Framework,” 2024. [Online]. Available: https://streamlit.io



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY