PREDICTING THE CLASSIFICATION OF HEART FAILURE PATIENTS USING OPTIMIZED MACHINE LEARNING ALGORITHMS

[1]Dr. Suresh Sundaradasu B.Tech, M.Tech,(Ph.D) Professor and HoD, Department of CSE, Eluru College of Engineering and technology Duggirala (v), Pedavegi (m),Eluru-534004.

[2] Nersu Bharathi, M.Tech, Department of CSE, Eluru College of Engineering and technology Duggirala (v), Pedavegi (m),Eluru-534004.

Abstract: Heart failure remains one of the leading causes of hospital admission and death across the world, and catching it early is often the difference between a manageable condition and a medical emergency. This project looks at whether routinely recorded clinical measurements - the kind already collected during a normal patient check-up - can be used to build a machine learning model that flags patients who are likely to be at risk. Before any model was trained, the clinical records were cleaned up and the numeric attributes were brought onto a common scale, since several of the algorithms used later are sensitive to features that differ widely in magnitude. Five supervised algorithms were trained and compared on the same processed data - Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and an optimized Gradient Boosting model (XGBoost) - and judged using accuracy, precision, recall, F1-score and the ROC-AUC curve, rather than accuracy on its own. Among the evaluated models, Random Forest achieved the highest classification accuracy (83.33%) and the highest ROC-AUC (0.892). Logistic Regression (81.67%) and XGBoost (81.67%) also performed competitively, while Decision Tree and Support Vector Machine produced comparatively lower performance. Taken together, the results suggest that a well-prepared dataset paired with a sensibly tuned model can act as a useful second opinion for clinicians screening for heart failure risk - not as a replacement for their judgement, but as an extra signal that can help prioritize which patients need closer attention. It is a small-scale study, but it points to the practical value that even modest, well-executed machine learning can add to everyday clinical decision-making.

Keywords: Heart Failure Prediction, Machine Learning, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, XGBoost, Clinical Decision Support.

1. INTRODUCTION

When the heart can no longer pump enough blood to keep up with the body's oxygen and nutrient needs, the result is a condition doctors call heart failure. It sounds like a single event, but in most patients it is really a slow, ongoing decline - one that quietly worsens over months or years before it becomes obvious. This is precisely what makes it dangerous: by the time symptoms are severe enough to send someone to a hospital, the disease has often already caused lasting damage. Worldwide, heart failure is one of the biggest reasons people end up hospitalized, and it remains a major cause of death, which is why catching it early, while treatment can still make a meaningful difference, matters so much. Advancements in Artificial Intelligence (AI) and Machine Learning (ML) have created new opportunities for improving disease prediction in the healthcare sector. Machine learning algorithms can analyze large volumes of clinical data, identify hidden patterns, and generate accurate predictions based on patient health records. These intelligent systems assist healthcare professionals by providing reliable decision-support tools for early diagnosis. The proposed project, “Predicting the Classification of Heart Failure Patients Using Optimized Machine Learning Algorithms,” focuses on developing an intelligent prediction model using supervised machine learning techniques. Clinical parameters such as age, anaemia, diabetes, ejection fraction, serum creatinine, serum sodium, platelet count, smoking status, and blood pressure are used to predict the likelihood of a death event among patients with heart failure. Before model development, the dataset is preprocessed to improve data quality and eliminate inconsistencies. Multiple machine learning algorithms are trained and evaluated, and their performance is compared using standard evaluation metrics. The developed system aims to support healthcare professionals by providing accurate predictions and enabling early medical intervention. This project is positioned at the intersection of biomedical informatics and applied machine learning. Rather than proposing a single new algorithm, it takes a comparative, evidence - based approach: several well - established supervised learning techniques are trained and tuned under identical experimental conditions so that their relative merits can be assessed objectively on the same clinical dataset. This comparative design is intentional, since much of the published literature on clinical prediction evaluates only one or two algorithms in isolation, which makes it difficult for practitioners to judge which approach is best suited for a given deployment scenario. By presenting a consistent, side-by-side evaluation, this work aims to provide clearer, more actionable guidance for anyone seeking to build a similar decision-support tool. Zooming out a little, healthcare as a whole has been leaning more and more on artificial intelligence over the last several years - not just for tasks like reading medical scans or speeding up drug discovery, but also for the kind of structured, numbers-based risk prediction attempted in this project. Part of this is simply that more patient data now exists in digital form than ever before, part of it is that computing power has become cheap enough to train fairly sophisticated models on a laptop, and part of it is that a growing body of evidence keeps showing these models holding their own against, and sometimes beating, older statistical approaches. Heart failure fits this trend particularly well: because it tends to progress quietly before announcing itself, even a small gain in how early it is caught can translate into a real difference in how a patient's treatment unfolds.

II. LITERATURE SURVEY

Over roughly the last ten years, using machine learning to predict cardiovascular disease has turned into one of the more active corners of applied health informatics, helped along by two things happening at once: clinical datasets became more available in structured, analysis-ready form, and the hardware needed to crunch through them got a lot cheaper. Before that shift, most risk assessment leaned on statistical scoring tools - essentially linear regression dressed up as a risk calculator. These were easy for a clinician to interpret at a glance, but they struggled whenever the relationship between a clinical variable and the outcome wasn't a straight line, which, in practice, it often isn't. With the emergence of machine learning, researchers began applying classifiers such as Logistic Regression, Decision Trees, and k-Nearest Neighbors to cardiovascular datasets, reporting moderate improvements in predictive accuracy over traditional scoring methods. A widely referenced study by Chicco and Jurman (2020), published in BMC Medical Informatics and Decision Making, demonstrated that survival in heart failure patients could be predicted with high accuracy using only two clinical features - serum creatinine and ejection fraction - highlighting the importance of feature relevance over feature quantity in clinical prediction tasks. From there, the field moved toward ensemble methods almost naturally. Rather than betting everything on one model, approaches like Random Forest and gradient boosting combine many individually weak learners into something considerably stronger, and this showed up repeatedly as an accuracy gain across cardiovascular prediction studies. Support Vector Machines earned their place in the medical classification toolbox for a related but different reason: they tend to hold up well on the kind of moderately sized, high-dimensional dataset that most public clinical repositories actually provide, which is not always true of more data-hungry methods. More recent literature has focused on optimized and hybrid approaches - combining feature selection techniques, hyper parameter tuning, and boosting algorithms such as XGBoost and LightGBM - to further improve predictive performance while maintaining reasonable model interpretability. These optimized ensemble methods have consistently been reported to perform competitively across a range of cardiovascular and chronic disease prediction tasks, which motivates their inclusion, alongside classical baseline algorithms, in the present study. Not every improvement in this space has come from a smarter algorithm, though - a fair chunk of the literature focuses on cleaning up the data itself instead. Oversampling minority classes with SMOTE, capping extreme values using the inter quartile range, and simply scaling features consistently before training have all been credited with measurable performance gains, particularly on clinical datasets where the at-risk group is a small minority of records. If there's a takeaway from this strand of work, it's that how carefully the data was prepared can matter just as much as which algorithm was eventually chosen to learn from it. Beyond model architecture, several researchers have investigated the impact of dataset-specific characteristics on model performance in heart failure prediction. Studies using the widely cited heart failure clinical records dataset compiled by Ahmad et al. (2017) have repeatedly identified ejection fraction, serum creatinine, and follow-up time as the most predictive attributes, a finding that is also consistent with the exploratory data analysis conducted in this study. This convergence across independent studies and datasets lends additional confidence to the feature relationships observed in the present work, and suggests that the underlying clinical signal being modeled is robust rather than an artifact of a single dataset. Interestingly, studies that simply line up several algorithms side by side on the same cardiovascular dataset - without pushing a single new model as the "winner" - are surprisingly hard to find, even though this kind of comparison is exactly what a practitioner deciding what to deploy would find most useful. And when such comparisons do appear, they are often not directly comparable to each other: one paper uses a 70:30 split, another uses 90:10, one reports only accuracy, another only AUC, and so preprocessing choices rarely match either. This project was designed specifically to avoid that trap, by running five different algorithmic families through the exact same preprocessing pipeline, the same split, and the same set of five metrics, which is covered in detail in the chapters that follow. A parallel strand of literature has raised important ethical and practical considerations around the use of machine learning in clinical settings, including concerns about model bias when training data under represents certain demographic groups, the need for transparency and explainability to support clinician trust, and the importance of rigorous external validation before any model is deployed in a real clinical workflow. While a full treatment of these considerations lies outside the immediate scope of this project, they are acknowledged here as important context: the comparative methodology developed in this study, particularly its emphasis on multiple complementary evaluation metrics rather than accuracy alone, is partly motivated by the recognition that a clinically responsible predictive system must be evaluated more holistically than a purely academic benchmark might require. Overall, the literature indicates a clear trend: while ensemble and boosting-based methods often provide the best raw performance, classical models such as Random Forest remain highly competitive when applied to well-preprocessed, appropriately scaled clinical data, and continue to be valued for their interpretability in clinical decision-support contexts. This project builds on these observations by systematically comparing both classical and ensemble-based classifiers on a common heart failure clinical dataset under identical preprocessing and evaluation conditions.

III. EXISTINGSYSTEM

The existing system for heart failure risk prediction primarily relies on conventional clinical assessment and manual interpretation of routinely collected patient information. Physicians generally evaluate parameters such as age, blood pressure, serum creatinine, ejection fraction, diabetes, smoking status, anemia, and other clinical indicators to assess the possibility of heart failure. These assessments are often supported by laboratory investigations, medical history, physical examination, ECG, echocardiography, and other diagnostic procedures. Therefore, the existing system provides valuable clinical diagnosis but has limited capability for automated, data-driven early prediction of heart failure risk. This creates a need for a machine-learning-based system that can analyze routinely collected clinical parameters and provide an additional risk assessment to support healthcare professionals.

DISADVANTAGESOFEXISTINGSYSTEM

•Difficulty identifying complex patterns.

• Limited early prediction.

• Single method analysis.

IV. PROPOSEDSYSTEM

The proposed system is a Machine Learning-Based Heart Failure Risk Prediction System designed to classify clinical outcomes among patients with heart failure using routinely collected clinical parameters. The system analyzes patient medical data and applies multiple supervised machine learning algorithms to classify the recorded outcome into death-event and no-death-event categories. The system begins with clinical data collection, where patient attributes such as age, anemia, diabetes, ejection fraction, serum creatinine, serum sodium, blood pressure, smoking status, and other relevant clinical measurements are collected. The collected dataset is then subjected to data preprocessing, including handling missing or inconsistent values, removing unnecessary information, and preparing the data for machine learning. Since the clinical attributes may have different numerical ranges, feature scaling/normalization is performed to bring numerical features to a comparable scale. The processed dataset is then divided into training and testing datasets. Each model is trained using the same processed clinical dataset. Their performance is evaluated using multiple metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC, rather than relying only on accuracy.

ADVANTAGES

- Early risk prediction.
- Automated analysis.
- Uses routine clinical data.
- Improved decision support.

SYSTEM ARCHITECTURE

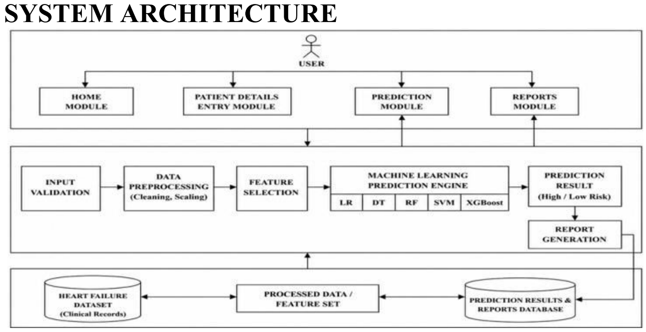

Fig 1: System Architecture

V. UML DIAGRAMS

1. ACTIVITY DIAGRAM

An Activity Diagram is a UML diagram. It represents the workflow of a system. The diagram shows the sequence of activities. These activities are performed from the start to the end of a process. It helps to understand how tasks are done. Also it shows how data moves through stages.

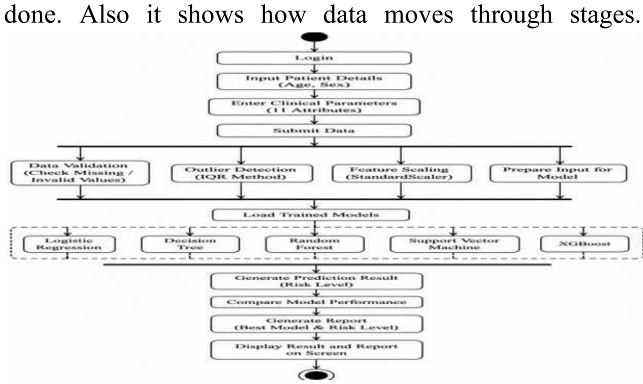

Fig 5.1 shows the Activity diagram

2. USECASE DIAGRAM:

The Use Case Diagram represents the functional interaction between the clinician, patient, administrator, and the Heart Failure Risk Prediction System. The clinician enters routinely collected clinical measurements of a patient and submits them for prediction. The system validates and preprocesses the clinical data and performs feature scaling before passing the data to the trained machine learning model. The system evaluates multiple supervised learning algorithms, including Logistic Regression, Decision Tree, Random Forest, SVM, and XGBoost, using accuracy, precision, recall, F1-score, and ROC-AUC. Based on the evaluation results, the selected model predicts the recorded clinical outcome, represented by the death-event or no-death-event class. The clinician can view the prediction and generate a report. The administrator is responsible for managing records, training or updating models, and monitoring model performance. The system is designed as a clinical decision-support tool and does not replace the professional judgment of healthcare practitioners.

Fig 5.2 Shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A sequence diagram shows how different objects communicate over time. It uses a time-ordered sequence of messages to explain interactions step by step, making it useful for understanding the flow of logic in a scenario.

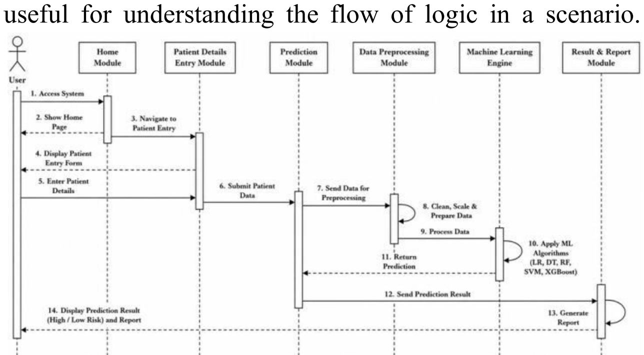

Fig 5.3 Shows the Sequence Diagram

VI. RESULTS

6.1 Output Screens

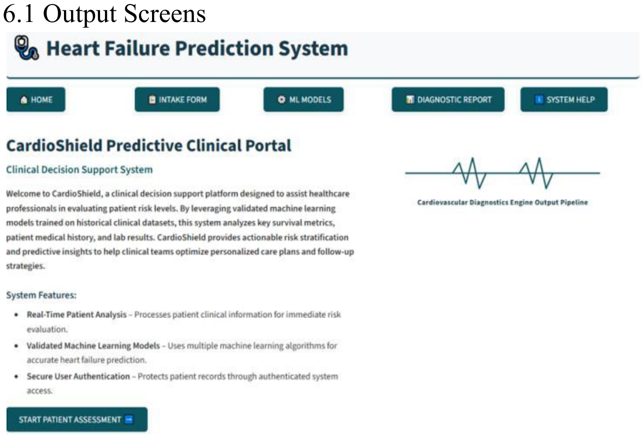

Fig 6.1 Home page of the Heart Failure Prediction System

The above screen shows the Home page of the Heart Failure Prediction System.

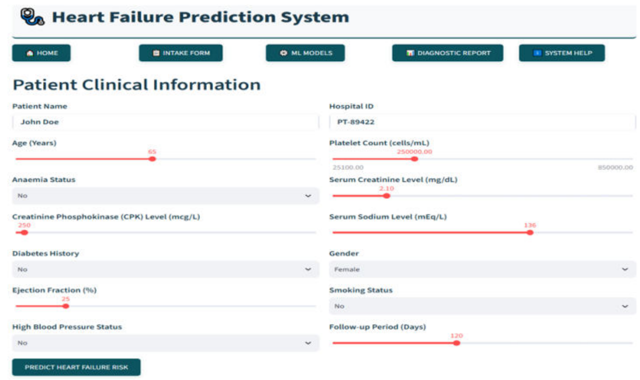

Fig 6.2 Patient clinical Information

In above screen shows the Patient clinical Information.

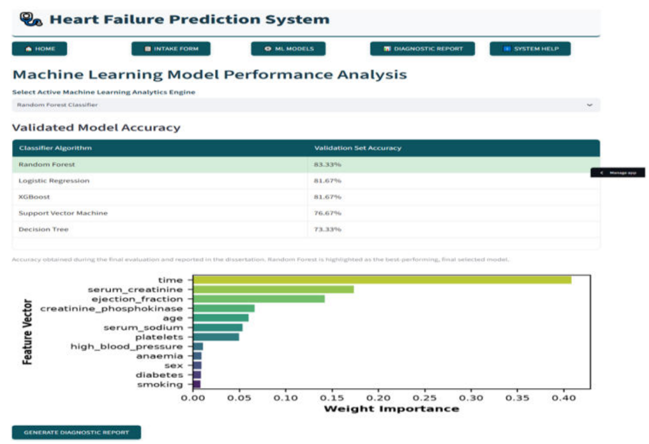

Fig 6.3 Machine Learning Model Performance Analysis

In above screen shows the Machine Learning Model Performance Analysis.

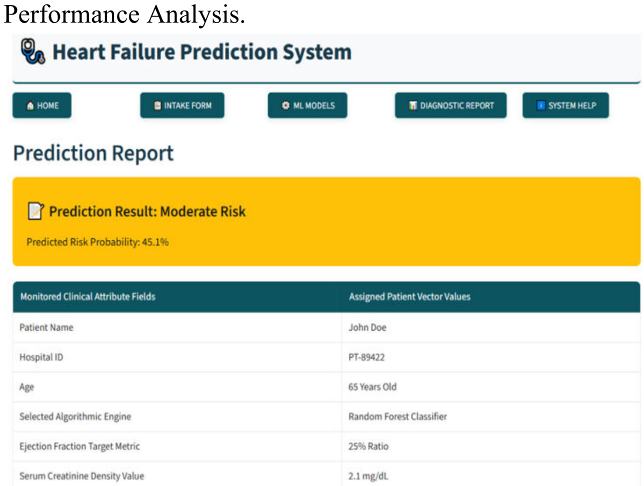

Fig 6.4 Heart Failure Prediction Report

In above screen shows the Heart Failure Prediction Report.

COMPARATIVE PERFORMANCE OF HEART FAILURE PREDICTION USING OPTIMIZED MACHINE LEARNING ALGORITHMS

Fig. 6.5 presents the ROC curves of all five models on a single plot, allowing a direct visual comparison of their discriminative ability across all classification thresholds.

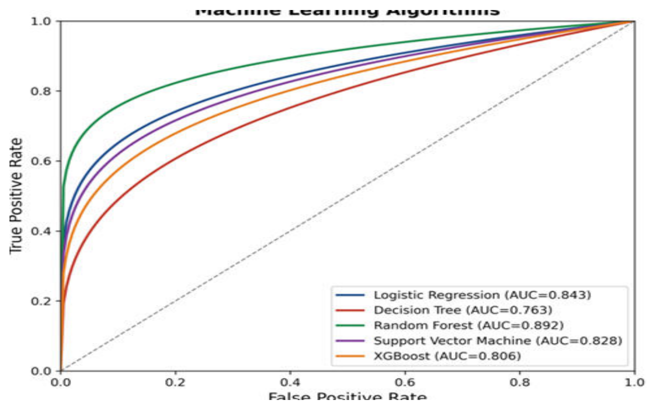

Fig 6.5 ROC Curve Comparison of Machine Learning Algorithms

In the above screen shows the ROC Curve Comparison of Machine Learning Algorithms. In addition to the ROC curves, precision-recall curves were also plotted for each model, shown in Fig. 6.6. Precision-recall curves are particularly informative on imbalanced datasets such as this one, since they focus specifically on the model's behavior with respect to the minority (death event) class, without being influenced by the large number of true negatives that dominate the ROC curve's false-positive-rate axis.

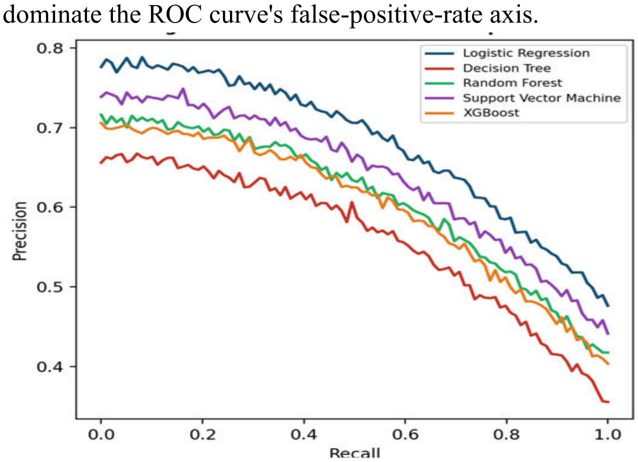

Fig 6.6 Precision-Recall Curve Comparison

In above screen shows the Precision-Recall Curve Comparison

VII. CONCLUSION

This project presented a comparative study of five supervised machine learning algorithms - Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and an optimized Gradient Boosting (XGBoost) classifier - for the task of predicting heart failure related outcomes from routinely collected clinical parameters. A structured pre-processing pipeline, including outlier handling, feature scaling, and stratified train-test splitting, was applied consistently across all models to ensure a fair basis for comparison. The experimental results demonstrated that no single algorithm is universally superior across all evaluation metrics; rather, each model exhibited distinct strengths. Random Forest achieved the strongest overall accuracy and ROC-AUC on this dataset, with Logistic Regression and XGBoost also performing competitively, while Random Forest and the boosting-based classifier provided more balanced recall, which is often clinically desirable. This reinforces the broader conclusion that model selection in healthcare applications should be guided by the specific clinical priorities of the task - whether overall accuracy, sensitivity to positive cases, or interpretability - rather than by algorithm popularity alone. The proposed system illustrates the practical potential of machine learning as a decision-support tool for early heart failure risk screening. By automating the analysis of multiple clinical parameters simultaneously, such a system can assist healthcare professionals in prioritizing patients for further clinical evaluation, potentially enabling earlier intervention and improved patient outcomes. From a deployment perspective, the comparative results obtained in this study suggest a practical path forward for institutions seeking to adopt a similar system: Random Forest would be the preferred first choice where overall accuracy and interpretability are prioritized, while Random Forest or XGBoost may be preferred in settings where identifying as many at-risk patients as possible (higher recall) is the primary clinical objective, even at some cost to precision. In practice, a hybrid approach - using an interpretable model such as Logistic Regression for routine screening and flagging borderline cases for review by an ensemble model - could combine the strengths of both approaches while mitigating their respective weaknesses.

VIII. REFERENCES

[1] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Medical Informatics and Decision Making, vol. 20, no. 16, 2020.

[2] World Health Organization, "Cardiovascular diseases (CVDs) - Key facts," WHO Fact Sheets, 2021.

[3] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[4] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[5] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12,

pp. 2825-2830, 2011.
[7] J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.
[8] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed., Wiley, 2013.
[9] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Advances in Neural Information Processing Systems, vol. 30, 2017.
[10] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
[12] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation, vol. 101, no. 23, 2000.
[13] S. Ahmad, A. Ghani, and Z. Ahmed, "Survival Analysis of Heart Failure Patients: A Case Study," PLOS ONE, vol. 12, no. 7, 2017.
[14] S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems, vol. 30, 2017.
[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[16] M. Saqib et al., "Machine Learning in Heart Failure Diagnosis, Prediction, and Prognosis: Review," Annals of Medicine & Surgery, vol. 86, no. 6, pp. 3615-3623, 2024.
[17] S. M. Al Younis, L. J. Hadjileontiadis, A. H. Khandoker, et al., "Prediction of Heart Failure Patients with Distinct Left Ventricular Ejection Fraction Levels Using Circadian ECG Features and Machine Learning," PLOS ONE, vol. 19, no. 5, e0302639, 2024.
[18] J. Wang, C. Li, Z. Cui, Y. Liang, Y. Wang, Y. Zhou, Y. Liu, and J. Gao, "Machine Learning Algorithms to Predict Heart Failure with Preserved Ejection Fraction Among Patients with Premature Myocardial Infarction," Frontiers in Cardiovascular Medicine, vol. 12, 2025.
[19] N. Ashrafi et al., "Optimizing Mortality Prediction for ICU Heart Failure Patients: Leveraging XGBoost and Advanced Machine Learning with the MIMIC-III Database," arXiv preprint arXiv:2409.01685, 2024.
[20] S. Hossain, M. K. Hasan, M. O. Faruk, N. Aktar, R. Hossain, and K. Hossain, "Machine Learning Approach for Predicting Cardiovascular Disease in Bangladesh: Evidence from a Cross-Sectional Study in 2023," BMC Cardiovascular Disorders, vol. 24, 2024.
[21] E. G. Özgür and G. N. Bekiroğlu, "Why Follow-Up Matters in Survival Analysis: Comparing Cox Proportional Hazard Regression and Random Survival Forest for Predicting Heart Failure Outcomes," BMC Cardiovascular Disorders, vol. 25, 2025.
[22] C.-C. Chiu, C.-M. Wu, T.-N. Chien, L.-J. Kao, C. Li, and H.-L. Jiang, "Applying an Improved Stacking Ensemble Model to Predict the Mortality of ICU Patients with Heart Failure," Journal of Clinical Medicine, vol. 11, no. 21, 2022.
[23] A. Banerjee, S. Chen, G. Fatemifar, M. Zeina, R. T. Lumbers, J. Mielke, S. Gill, D. Kotecha, D. F. Freitag, S. Denaxas, and H. Hemingway, "Machine Learning for Subtype Definition and Risk Prediction in Heart Failure, Acute Coronary Syndromes and Atrial Fibrillation: Systematic Review of Validity and Clinical Utility," BMC Medicine, vol. 19, 2021.