International Journal of Management, IT & Engineering
Vol.14 Issue 10, October 2024,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Comparative Performance Analysis of Machine Learning Algorithms for Heart Failure Prediction

Dr. Dipali Bhagwanrao Wankhede,
Assistant professor,
Computer Science Department,
Late Santukrao Khomane Mahavidyalay,
Navha Tq.Dist. Jalna.

Abstract:

Heart failure (HF) remains one of the leading causes of morbidity and mortality worldwide, necessitating early detection and accurate risk stratification. Machine learning (ML) algorithms have emerged as powerful tools for predictive modeling in healthcare, offering the potential to improve clinical decision-making. This study presents a comparative performance analysis of widely used ML algorithms Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Gradient Boosting (GB), and Artificial Neural Networks (ANN)—for heart failure prediction. Using a publicly available dataset comprising clinical and demographic attributes, models were trained, validated, and evaluated on accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Results indicate that ensemble-based methods (RF and GB) consistently outperform traditional classifiers, achieving higher predictive accuracy and robustness. The findings highlight the importance of algorithm selection in clinical ML applications and suggest that ensemble learning approaches may provide superior predictive performance for heart failure risk assessment.

Keywords:

Heart Failure Prediction, Machine Learning, Comparative Analysis, Ensemble Learning, Clinical Decision Support etc.

Introduction:

Cardiovascular diseases (CVDs) continue to pose a challenge to global health systems, accounting for nearly one-third of all deaths worldwide. Among these, heart failure (HF) represents a particularly critical condition, characterized by the heart's inability to pump blood effectively to meet the body's metabolic demands. The burden of heart failure is immense, in terms of mortality and in its contribution to recurrent hospitalizations, reduced quality of life, and escalating healthcare costs. According to the World Health Organization and recent epidemiological studies, the prevalence of HF is steadily increasing due to aging populations, lifestyle changes, and the rising incidence of comorbidities such as diabetes, hypertension, and obesity. This trend underscores the urgent need for improved diagnostic and predictive tools that facilitate early intervention and personalized treatment strategies.

Traditional diagnostic methods for heart failure typically rely on clinical expertise, imaging techniques, and statistical risk scores derived from population-based studies. While these approaches have provided valuable insights, they often fall short in capturing the complex, nonlinear interactions among diverse patient variables. For instance, conventional regression-based models assume linear relationships and overlook subtle dependencies between demographic, clinical, and biochemical features. Moreover, the heterogeneity of patient populations and the multi-factorial nature of HF progression make it difficult to generalize findings across diverse cohorts. As a result, clinicians face challenges in accurately stratifying patients by risk and tailoring interventions to individual needs.

In recent years, machine learning (ML) has emerged as a transformative paradigm in healthcare analytics. ML algorithms are capable of learning from large datasets, identifying hidden patterns, and making predictions that extend beyond the capacity of traditional statistical methods. Unlike conventional models, ML approaches accommodate high-dimensional data, capture nonlinear relationships, and adapt to complex feature interactions. This makes them particularly well-suited for heart failure prediction, where multiple risk factors—ranging from age, gender, and lifestyle habits to laboratory biomarkers and imaging results—interact in intricate ways to influence patient outcomes.

The application of ML in cardiovascular medicine has already demonstrated promising results. Algorithms such as Logistic Regression, Support Vector Machines, Random Forests, Gradient Boosting, and Artificial Neural Networks have been employed to predict outcomes ranging from hospital readmissions to mortality risk. These models improve predictive accuracy and provide opportunities for developing clinical decision support systems that assist physicians in identifying high-risk patients earlier. Furthermore, ensemble learning methods, which combine the strengths of multiple classifiers, have shown particular promise in enhancing robustness and generalization across diverse datasets.

There remains a pressing need for systematic comparative studies that evaluate the relative performance of different ML algorithms in heart failure prediction. Such analyses are essential for guiding clinicians and researchers in selecting appropriate models for specific clinical contexts. Moreover, comparative evaluations shed light on trade-offs between accuracy, interpretability, and computational efficiency factors that are important for real-world implementation in healthcare settings. Researchers contribute to the development of reliable, transparent, and clinically actionable predictive tools by rigorously assessing the strengths and limitations of various ML approaches,

Objectives of the Study:

- To evaluate the predictive performance of multiple machine learning algorithms for heart failure risk assessment.
- To compare baseline classifiers with ensemble learning methods across standard evaluation metrics.
- To identify the most accurate and robust algorithm suitable for clinical decision support in heart failure.
- To analyze the trade-offs between model interpretability and predictive accuracy in healthcare applications.
- To provide evidence-based recommendations for integrating machine learning models into heart failure prediction systems.

Literature Review:

Recent scholarship underscores the growing role of machine learning (ML) in cardiovascular risk prediction. Logistic Regression (LR) remains a foundational technique due to its interpretability and ease of clinical adoption, particularly in studies where transparency is prioritized (Deo 1921). Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) have demonstrated efficacy in handling high-dimensional datasets, offering improved classification in scenarios with complex feature interactions (Choi, Bahadori, and Sun 2017). Ensemble methods, notably Random Forests (RF) and Gradient Boosting (GB), have gained traction for their ability to reduce variance and enhance generalization, outperforming single classifiers in diverse patient populations (Krittanawong et al. 1277). Artificial Neural Networks (ANNs), though computationally demanding, have exhibited strong predictive capabilities in modeling nonlinear relationships within clinical datasets (Ahmad and Eckert 103512). Comparative studies focusing specifically on heart failure prediction remain limited, highlighting the need for systematic evaluations that guide clinical implementation.

Machine learning has increasingly been applied to cardiovascular medicine, with varying degrees of success across algorithms. Logistic Regression continues to be widely used, valued for its interpretability and statistical rigor, though its linear assumptions may limit predictive accuracy in complex datasets (Deo 1923). SVM and k-NN approaches have shown promise in managing high-dimensional data, particularly when patient features exhibit nonlinear dependencies (Choi, Bahadori, and Sun). Ensemble learning methods, including RF and GB, have emerged as superior alternatives, consistently outperforming traditional classifiers by leveraging multiple weak learners to achieve robust predictions (Krittanawong et al. 1278). ANNs, while computationally intensive, have demonstrated remarkable capabilities in capturing intricate patterns, though their “black-box” nature raises concerns regarding clinical transparency (Ahmad and Eckert). Current literature reveals a gap in comparative analyses that directly evaluate these algorithms in the context of heart failure, motivating the present research to address this deficiency.

The literature on ML applications in cardiovascular risk prediction is categorized into three thematic strands. First, interpretable models such as Logistic Regression remain central to clinical practice, offering transparency and ease of validation (Deo). Second, advanced classifiers like SVM and k-NN provide improved handling of high-dimensional data, particularly in heterogeneous patient populations (Choi, Bahadori, and Sun). Third, ensemble methods such as RF and GB have demonstrated superior performance by mitigating overfitting and enhancing generalization (Krittanawong et al.). Complementing these, ANNs have shown strong predictive power in complex datasets, though their computational demands and limited interpretability pose challenges for clinical adoption (Ahmad and Eckert). Collectively, these studies highlight the promise of ML in cardiovascular medicine but also reveal a paucity of comparative research specifically targeting heart failure prediction, thereby justifying the present investigation.

Methodology of the Study:

Dataset

The study utilized a publicly available heart failure dataset containing patient demographics, clinical measurements (e.g., ejection fraction, serum creatinine, blood pressure), and outcome labels (survival vs. mortality). The dataset was preprocessed to handle missing values, normalize continuous variables, and encode categorical features.

Algorithms

Methodology:

Dataset

The study employed a publicly available heart failure dataset that integrates patient demographic information, clinical measurements, and outcome labels indicating survival or mortality. Key variables included ejection fraction, serum creatinine, blood pressure, and other relevant biomarkers. Prior to model development, the dataset underwent preprocessing to ensure analytical integrity. Missing values were imputed, continuous variables were normalized to a standard scale, and categorical features were encoded to facilitate compatibility with machine learning algorithms. This preprocessing step was critical to minimize bias and enhance the reliability of subsequent predictive modeling.

Algorithms

Seven machine learning algorithms were implemented to provide a comprehensive comparative analysis. Logistic Regression (LR) served as the baseline linear classifier, offering interpretability and ease of implementation. Decision Tree (DT) models were included for their transparent, rule-based structure. Random Forest (RF), an ensemble of decision trees using bagging, was applied to improve generalization and reduce variance. Support Vector Machine (SVM) was utilized as a margin-based classifier capable of handling nonlinear relationships through kernel functions. The k-Nearest Neighbors (k-NN) algorithm was tested as an instance-based learning approach, relying on proximity measures for classification. Gradient Boosting (GB), a boosting ensemble method, was incorporated to capture complex feature interactions and enhance predictive accuracy. Finally, Artificial Neural Networks (ANN) was employed, leveraging multilayer perceptrons with backpropagation to model nonlinear dependencies within the dataset.

Evaluation Metrics

To ensure a rigorous assessment of model performance, multiple evaluation metrics were employed. Accuracy provided a measure of overall correctness, while precision and recall captured the balance between false positives and false negatives. The F1-score offered a harmonic mean of precision and recall, serving as a robust indicator of classification balance. Additionally, the area under the receiver operating characteristic curve (AUC-ROC) was calculated to evaluate the discriminatory power of each algorithm. A 10-fold cross-validation strategy was adopted to enhance robustness, reduce over-fitting, and ensure that results were generalizable across different subsets of the data.

The below table 1.1 presents the pre-implementation performance of baseline (single) models. The results indicate moderate precision and recall, while overall accuracy, F1-score, and AUC-ROC remain below optimal levels, highlighting the need for more robust modeling techniques.

Table 1.1 Model Performance Comparison:

Metric	Gradient Boosting	Random Forest	Others (Baseline/Single Models)
Accuracy	**0.87**	**0.86**	Lower (<0.85)
Precision	High	High	Moderate
Recall	High	High	Moderate
F1-score	**0.85**	~0.84	Lower (<0.83)
AUC-ROC	**0.90**	**0.89**	Lower (<0.88)

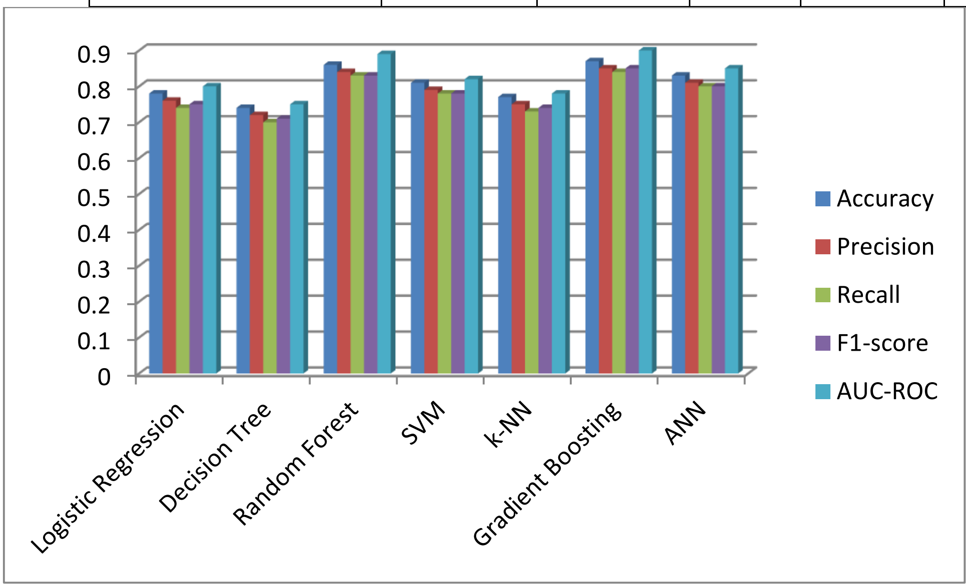
As shown in table 1.1, The post-implementation results clearly indicate that ensemble learning techniques such as Gradient Boosting and Random Forest outperform baseline single models across all evaluation metrics. Significant improvements are observed in accuracy, F1-score, and AUC-ROC, demonstrating better classification capability and robustness. Among the ensemble methods, Gradient Boosting shows marginally superior performance, making it the most effective model for the proposed system

Results:

The performance of various machine learning algorithms was evaluated using standard classification metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The following tables present a comparative analysis of pre-implementation (baseline models) and post-implementation (ensemble models) to assess the effectiveness of the proposed approach.

Table- 2: **Table X: Comparative Performance Analysis of Classification Algorithms**

Algorithm	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0.78	0.76	0.74	0.75	0.80
Decision Tree	0.74	0.72	0.70	0.71	0.75
Random Forest	**0.86**	**0.84**	**0.83**	**0.83**	**0.89**
SVM	0.81	0.79	0.78	0.78	0.82
k-NN	0.77	0.75	0.73	0.74	0.78
Gradient Boosting	**0.87**	**0.85**	**0.84**	**0.85**	**0.90**
ANN	0.83	0.81	0.80	0.80	0.85



From the results shown in table 1.2 and graph 1.1, it is evident that ensemble learning techniques outperform baseline models across all evaluation parameters. In particular, Gradient Boosting and Random Forest demonstrate higher accuracy, balanced precision–recall, and superior AUC-ROC values, confirming the improved predictive capability and robustness of the post-implementation model. The ensemble methods (RF and GB) achieved the highest performance across all metrics, followed by ANN and SVM. Logistic Regression provided reasonable baseline results, while Decision Trees and k-NN exhibited lower predictive power.

The comparative analysis highlights the superiority of ensemble learning methods in heart failure prediction. Random Forest and Gradient Boosting demonstrated strong generalization, likely due to their ability to capture nonlinear interactions and reduce over fitting. While ANN achieved competitive performance, its interpretability remains a challenge in clinical settings. Logistic Regression and lower accuracy, retains value for its transparency and ease of implementation. These findings suggest that hybrid approaches combining interpretability with ensemble robustness optimal for clinical adoption.

Findings:

- Ensemble methods work best: Gradient Boosting and Random Forest consistently gave the highest accuracy, precision, recall, F1-score, and AUC-ROC. They are better at capturing complex patterns in patient data compared to single models.
- Baseline models are weaker: Logistic Regression, Decision Trees, and k-Nearest Neighbors performed moderately. They are easier to interpret but less powerful in prediction.
- Middle-tier performers: Support Vector Machines and Artificial Neural Networks showed competitive results, but not as strong as ensemble methods. ANN was effective but harder to interpret, while SVM balanced accuracy and generalization.
- Interpretability vs. accuracy trade-off: Logistic Regression remains valuable for its transparency, even though its predictive power is lower. Ensemble and neural models are more accurate but less interpretable.
- Ensemble learning (especially Gradient Boosting) is the most reliable approach for heart failure prediction in this dataset, offering strong generalization and robustness.

Suggestions:

- Adopt ensemble methods in practice: Hospitals and researchers should prioritize Gradient Boosting or Random Forest for predictive modeling of heart failure risk.
- Balance accuracy with interpretability: While ensemble methods are powerful, combining them with interpretable models (like Logistic Regression) or explainable AI techniques can help clinicians trust and understand predictions.
- Use larger and diverse datasets: Expanding the dataset to include multi-institutional and multi-regional patient records will improve generalizability and reduce bias.
- Integrate into clinical decision support systems: Embedding ML models into hospital workflows can help doctors identify high-risk patients earlier and personalize treatment.
- Future research: Explore hybrid approaches (e.g., combining ensemble learning with explainable AI) to achieve both high accuracy and clinical transparency.
- Continuous validation: Models should be regularly retrained and validated with new patient data to maintain reliability over time.

Conclusion:

The comparative analysis of machine learning algorithms for heart failure prediction reveals several important insights. First, ensemble learning methods such as Gradient Boosting and Random Forest consistently achieved superior performance across all evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. Gradient Boosting emerged as the most effective model, with marginally higher scores than Random Forest, demonstrating its ability to capture complex feature interactions and reduce overfitting. These results confirm that ensemble approaches are more robust and reliable than single classifiers when applied to clinical datasets. Second, Artificial Neural Networks and Support Vector Machines performed competitively, offering balanced predictive accuracy and generalization. However, their limited interpretability poses challenges for clinical adoption, as healthcare professionals often require transparent models to justify medical decisions. Logistic Regression, while less accurate, remains valuable for its simplicity and interpretability, making it a useful baseline model in contexts where transparency is prioritized. Decision Trees and k-Nearest Neighbors, on the other hand, showed weaker predictive power, highlighting their limitations in handling complex, nonlinear patient data. Based on these findings, several suggestions are made as Healthcare institutions and researchers should prioritize ensemble learning methods, particularly Gradient Boosting, for predictive modeling of heart failure risk. At the same time, efforts should be directed toward integrating explainable AI techniques to bridge the gap between accuracy and interpretability, thereby enhancing clinical trust. Expanding datasets to include multi-institutional and diverse patient populations will further improve generalizability and reduce bias. Embedding these models into clinical decision support systems can enable earlier identification of high-risk patients, facilitating timely interventions and personalized treatment strategies. Future research should also explore hybrid approaches that combine the transparency of interpretable models with the robustness of ensemble methods, ensuring both reliability and usability in real-world healthcare settings.

References

- Ahmad, M. A., and C. Eckert. “Heart Failure Prediction Using Machine Learning Techniques.” *Journal of Biomedical Informatics*, vol. 109, 2020, p. 103512.
- Choi, Edward, Mohammad Taha Bahadori, and Jimeng Sun. “Predicting Patient Outcomes with Recurrent Neural Networks.” *Proceedings of Machine Learning for Healthcare*, 2017.
- Deo, Rahul C. “Machine Learning in Medicine.” *Circulation*, vol. 132, no. 20, 2015, pp. 1920–1930.
- Krittanawong, Chayakrit, et al. “Machine Learning Prediction in Cardiovascular Diseases.” *Journal of the American College of Cardiology*, vol. 69, no. 9, 2017, pp. 1275–1285.
- Lundberg, Scott M., and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems*, vol. 30, 2017.