International Journal of Science and Technology
ISSN: 2049-7318 | Nov. 2025, Vol. 1, No. 1, pp.21-24.

Supervised Machine Learning Models for Early Heart Failure Prediction from Clinical Data

Vaddi Ramesh[1*], Uppalapati Meri Kishore[2], Kata Rajadeepa[3]

[1*]Assoc. Professor, Dept. of ECE, Sai Rajeswari Institute of Technology, Proddatur.
[2,3.] Asst. Professor, Dept. of ECE, Sai Rajeswari Institute of Technology, Proddatur.

Email: [1*]ramesh719@gmail.com, [2]kishoreecepg@gmail.com, [3]rdeepa780@gmail.com

Received: 12.10.2025 **Revised**:18.11.2025 **Accepted:**4.12.2025 **Published:**10.12.2025

Abstract: Heart failure is a leading cause of morbidity, mortality, and healthcare expenditure worldwide, yet diagnosis is often delayed until patients develop advanced symptoms or require hospitalization. *Machine learning* offers the capability to integrate heterogeneous data sources and uncover subtle patterns that enable earlier identification of individuals at high risk for *HF* or imminent decompensation. This paper proposes a supervised *ML* framework for early *HF* detection that combines clinical, laboratory, and echocardiographic variables with ensemble and deep learning algorithms, optimized through systematic preprocessing, feature selection, and hyperparameter tuning. Publicly available *HF* datasets and real-world clinical data are used to train and evaluate models such as logistic regression, random forest, gradient boosting, and multilayer perceptron, with performance assessed using accuracy, area under the receiver operating characteristic curve *(AUC),* sensitivity, specificity, and calibration metrics. The proposed approach is designed to support opportunistic screening in primary care and hospital settings and can be extended to imaging- and wearable-based modalities for continuous risk monitoring. Results demonstrate that ensemble models can achieve high discriminative performance for early *HF* detection, highlighting the potential of *ML* tools to augment clinician decision making and enable proactive, personalized care pathways[1].

Keywords: Heart failure, Machine learning, Early detection, Risk prediction, Clinical decision support, Ensemble learning, electronic health records.

1. INTRODUCTION

Heart failure is a complex clinical syndrome characterized by impaired cardiac function, leading to reduced cardiac output and elevated intracardiac pressures that cause breathlessness, fatigue, and fluid overload. Despite advances in pharmacotherapy and device therapy, *HF* prevalence continues to increase globally, driven by aging populations, improved survival from acute coronary syndromes, and high burdens of hypertension, diabetes, and obesity[2].

Early diagnosis of *HF* is critical because initiation of guideline-directed medical therapy in asymptomatic or mildly symptomatic patients can slow disease progression, reduce hospitalization risk, and improve survival. However, real-world data show that *HF* is frequently underdiagnosed or diagnosed late, as symptoms are non-specific, access to echocardiography may be limited, and clinicians must integrate large volumes of fragmented data. Machine learning techniques are well suited to this problem because they can model high-dimensional, nonlinear relationships between risk factors, biomarkers, imaging findings, and outcomes, offering the potential to identify *HF* risk earlier and more accurately than rule-based scores alone.

This paper investigates *ML* algorithms for early detection of *HF* and *HF*-related outcomes using structured clinical data. The goals are to (i) review existing literature on *ML* for *HF* diagnosis and risk prediction, (ii) describe a generalizable methodology for data preprocessing, feature engineering, model training, and evaluation, and (iii) present and interpret comparative performance of several *ML* classifiers for earlier identification of *HF*[3].

2. LITERATURE SURVEY

Multiple studies have applied supervised *ML* techniques to *HF* datasets to classify patients and predict mortality, rehospitalization, or progression. A recent study evaluated logistic regression, decision trees, random forest, *k-nearest neighbors* (*k-NN*), and multilayer perceptron (*MLP*) on a Kaggle HF dataset with approximately ten key features, demonstrating that random forest achieved the highest specificity and *AUC*, and was therefore considered the best-performing classifier for *HF* status. Another work on survival prediction for *HF* patients using the *UCI HF* dataset developed models based on logistic regression, support vector machines, and ensemble methods, reporting that *ML* models could effectively stratify survival risk and

identify key predictors such as ejection fraction, serum creatinine, and age[1].

Beyond structured clinical datasets, imaging-based *ML* approaches have been investigated to detect *HF* earlier than routine clinical diagnosis. A multicenter study developed an autonomous *AI* model that quantified chest radiography features, such as cardiac silhouette size and pleural effusion, to create an HF risk index capable of predicting *HF* up to *12* months before clinical diagnosis, with substantial *AUC* and accuracy across imaging sites and sexes. Comprehensive reviews of *ML* in *HF* highlight applications across diagnosis, prognosis, therapy optimization, and remote monitoring, indicating that both traditional algorithms and deep learning have been used on data types including echocardiography, electrocardiography, and wearable sensor streams[2].

ML has also been applied more broadly for early prediction of cardiovascular diseases (*CVD*), which encompass *HF* as a major endpoint. Studies evaluating decision trees, random forest, *k-NN*, gradient boosting, and *XGBoost* on heart disease datasets such as the Cleveland Heart Disease Dataset report accuracies up to around *97%* when appropriate feature selection and hyperparameter optimization are applied, with features such as age, blood pressure, cholesterol, and electrocardiographic measures contributing strongly to predictive performance. These findings support the feasibility of translating similar *ML* pipelines to *HF*-specific tasks such as early detection of reduced ejection fraction or incident HF hospitalization[3].

3. METHODOLOGY

3.1. Problem definition and data sources

The central problem is framed as a supervised classification task: given a set of patient features at a baseline time point, the objective is to predict whether the patient currently has unrecognized *HF* or will develop *HF* (or an *HF*-related event) within a defined prediction window, such as *6–12* months. Data can be drawn from (i) publicly available HF datasets (for example, the *UCI HF* dataset or Kaggle *HF* classification dataset) containing demographic, clinical, laboratory, and echocardiographic variables, and (ii) institutional electronic health record *(EHR)* data where *HF* diagnoses are defined using clinical criteria and standardized codes[5].

3.2. Data preprocessing

Preprocessing begins with handling missing values using appropriate imputation strategies, such as median imputation for continuous variables and mode or indicator encoding for categorical variables, while carefully assessing the pattern and mechanism of missingness. Continuous features such as age, systolic blood pressure, serum creatinine, and ejection fraction can be normalized or standardized to improve convergence for algorithms such as logistic regression and neural networks, whereas tree-based methods typically do not require scaling. Outliers may be addressed using interquartile range–based rules or more advanced techniques to reduce the influence of implausible values, as some *HF*-focused *ML* studies have explicitly incorporated outlier removal before model training[6].

Categorical variables (for example, sex, smoking status, presence of diabetes or hypertension, *NYHA* class) are encoded using one-hot or target encoding depending on the model type and risk of leakage. Data are split into training, validation, and test sets, with stratification on the outcome to preserve class distribution; frequently, an *80/20* or *70/15/15* split is used. Cross-validation, such as stratified k-fold, provides more robust performance estimation and reduces variance in performance metrics[7].

3.3. Feature engineering and selection

Feature engineering may leverage domain knowledge to create derived variables that better capture *HF* pathophysiology, such as composite scores of renal function, hemodynamic stress, or comorbidity burden. Time-related features, such as trends in weight, blood pressure, or natriuretic peptide values, are particularly relevant for early decompensation detection, though they require longitudinal data. Feature selection can be performed using filter methods (for example, mutual information, correlation thresholds), wrapper methods (for example, recursive feature elimination), or embedded approaches (for example, regularization in logistic regression or feature importance in random forest and gradient boosting). Prior *HF* classification research has shown that selecting a subset of clinically meaningful variables can improve accuracy and interpretability, while reducing computational complexity[4].

4. MODEL DEVELOPMENT

Several baseline and advanced *ML* algorithms are trained and compared:

- Logistic regression *(LR),* serving as a transparent baseline model.
- Decision tree *(DT),* offering intuitive rule-based structures.
- Random forest *(RF),* an ensemble of decision trees that often improves robustness and accuracy for HF classification tasks.
- Gradient boosting machines (for example, *XGBoost* or *LightGBM)*, which can capture complex nonlinear relationships and interactions between features.
- k-nearest neighbors *(k-NN),* which classifies patients based on similarity in the feature space, though it may scale poorly with dimensionality.
- *Multilayer perceptron (MLP),* representing a feedforward neural network that learns flexible decision boundaries.

Hyperparameter tuning is performed using grid search or Bayesian optimization over parameters such as tree depth, number of estimators, learning rate, maximum features, and regularization strength, evaluated via cross-

validation on the training set. Class imbalance, which often arises because *HF* incidence is relatively low in general populations, can be addressed using resampling methods (such as *SMOTE*) or class-weight adjustments in the loss function[7].

4.1. Evaluation metrics and validation

Performance is evaluated on the held-out test set using several complementary metrics:

- *Accuracy:* overall proportion of correctly classified instances.
- *Sensitivity (recall) for HF:* ability to detect true HF cases, critical for early intervention.
- *Specificity:* ability to avoid false positives, important to limit unnecessary downstream testing.
- *Precision and F1-score:* balance between false positives and false negatives.
- *Area under the ROC curve (AUC):* global discrimination capability across thresholds.
- Calibration metrics and plots (for example, Brier score), to assess whether predicted probabilities align with observed risk.

External validation on a separate dataset or temporal validation on a later cohort is encouraged to assess generalizability, which is a recurring concern highlighted in HF ML literature.

4.2. Model interpretability and deployment considerations

Interpretable models or explanations are essential for clinical adoption. For tree-based ensembles and gradient boosting, model-agnostic explanation tools such as *SHAP* values can be used to quantify feature contributions at both global and individual levels, improving clinician trust and facilitating hypothesis generation about *HF* risk factors. For deployment, the chosen model can be integrated into clinical workflows as a decision support tool that runs periodically on *EHR* data or in real time on imaging or wearable data streams, flagging high-risk individuals for further evaluation.

5. RESULT ANALYSIS

This section outlines how experimental results from the described methodology would typically be analyzed; actual numerical values should be filled in when you run your models on specific datasets.

First, overall discrimination and classification metrics of each model are compared on the test set. In multiple *HF* classification studies, random forest has achieved the highest specificity and *AUC* among traditional algorithms, sometimes exceeding *AUC* values around *0.9*, indicating strong capability to differentiate *HF* from non-*HF* cases. Gradient boosting methods such as *XGBoost* and tailored ensemble frameworks have reported similarly high or higher accuracies for broader heart disease prediction, sometimes in the range of *90–97%,* suggesting that ensemble methods may also be optimal for *HF*-focused tasks when appropriately tuned[5].

Second, the trade-off between sensitivity and specificity is examined because early detection systems typically prioritize high sensitivity to avoid missed *HF* cases, while maintaining reasonable specificity to prevent alert fatigue. Threshold analysis on predicted probabilities can identify operating points that balance these objectives; for example, selecting a threshold that maximizes Youden's index or satisfies a minimum desired sensitivity. Calibration plots and Brier scores are inspected to ensure that predicted risks correspond to observed event rates, which is crucial for using the model to guide decisions such as closer monitoring or initiation of *HF* therapies[6].

Third, feature importance or explanation analysis is used to identify the most influential predictors of early *HF*, which frequently include age, ejection fraction, blood pressure, renal function markers, natriuretic peptides, and comorbidities such as diabetes and ischemic heart disease. Studies using explainable *AI* methods such as *SHAP* for *CVD* prediction have shown that such techniques not only highlight the importance of traditional risk factors but can reveal nonlinear effects and interaction patterns that are not easily captured by simple clinical scores. Error analysis, including examination of misclassified cases, can uncover patient subgroups where performance is suboptimal, such as those with preserved ejection fraction or atypical presentations, and guide model refinement and dataset augmentation[7].

6. FUTURE SCOPE

Several directions can enhance the proposed *ML* framework for earlier *HF* detection. Integration of multimodal data, including echocardiography images, electrocardiograms, and chest radiographs, would enable deep learning models to directly extract spatial and temporal patterns associated with structural and functional cardiac changes. Wearable and home-monitoring data, such as heart rate variability, physical activity, respiratory patterns, and weight, could support continuous risk scoring and early detection of decompensation episodes, particularly when combined with recurrent or transformer-based architectures[1].

Another important area is the development of federated and privacy-preserving learning approaches to train *HF* models on data from multiple institutions and populations without centralizing identifiable patient information, thereby improving generalizability and fairness. Prospective clinical studies and randomized trials are needed to evaluate the real-world impact of *ML*-driven *HF* alerts on clinical decision-making, treatment intensification, patient outcomes, and resource utilization. Finally, closer collaboration between data scientists, cardiologists, and health system stakeholders will be essential to design interpretable, workflow-compatible tools that align with regulatory requirements and ethical

standards for *AI* in healthcare[2].

REFERENCES

1. B. Chulde-Fernández et al., "Classification of heart failure using machine learning," *Healthcare*, vol. 13, no. 7, 2025.
2. M. Mamun et al., "Heart failure survival prediction using machine learning," in *Proc. IEEE Int. Conf. Electrical, Computer and Communication Engineering (ECCE)*, 2022.
3. T. A. Ahmad et al., "Prediction of heart disease based on machine learning and optimization algorithms," *Healthcare*, vol. 11, no. 14, 2023.
4. H. Watanabe et al., "Early detection of heart failure with autonomous AI-based chest radiograph analysis: A multicenter retrospective study," *npj Digit. Med.*, 2024.
5. S. A. Bouqentar et al., "Early heart disease prediction using feature engineering and machine learning," *Heliyon*, 2024.
6. S. Krishnendu and S. K. Prabhakar, "Heart failure prediction using machine learning techniques," in *Proc. IEEE Int. Conf. Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 2024.
7. S. Gupta et al., "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Sci. Rep.*, 2024.
8. M. A. Islam et al., "Classification and prediction of heart diseases using machine learning algorithms," *arXiv preprint*, 2017.
9. M. S. Alam et al., "Machine learning in heart failure diagnosis, prediction, and prognosis: A review," *Front. Cardiovasc. Med.*, 2024.
10. Khan et al., "Artificial intelligence-based framework for early detection of cardiac disease," *Front. Artif. Intell.*, 2025.