Predicting the classification of Heart failure patients using optimizing machine learning algorithms

[1]Naga V Vara Prasad Mella
[1]M.Tech, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
varaprasadmella@gmail.com

[2]Mr. B. V. Ram Kumar
[2]Assistant Professor, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
bvrk.bonam@gmail.com

Abstract*— Heart failure (HF) remains a leading cause of morbidity and mortality worldwide, and early, accurate classification of patient risk is essential for timely clinical intervention. This paper presents an optimized machine learning framework for predicting the classification of heart failure patients using structured clinical record data. The proposed approach integrates data preprocessing, feature selection, and hyperparameter optimization applied to multiple classifiers, including Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Logistic Regression. Key clinical features analyzed include age, ejection fraction, serum creatinine, serum sodium, platelet count, and comorbidity indicators such as diabetes, anaemia, and hypertension. Hyperparameters are tuned using Grid Search and Bayesian optimization to maximize predictive performance while reducing overfitting. Models are evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. Experimental analysis indicates that the optimized XGBoost model achieves the strongest overall classification performance, with an accuracy of approximately 96%, outperforming non-optimized baseline models and demonstrating the practical value of hyperparameter optimization for clinical decision support in heart failure prediction.

Keywords*— Heart Failure, Machine Learning, Classification, Hyperparameter Optimization, XGBoost, Random Forest, Clinical Prediction, Feature Selection, Explainable AI, Healthcare Analytics.

I. INTRODUCTION

Cardiovascular diseases remain the leading cause of death globally, with heart failure (HF) representing one of the most severe and costly manifestations of this disease group [1], [2]. HF occurs when the heart is unable to pump sufficient blood to meet the metabolic needs of the body, leading to progressive clinical deterioration if not identified and managed early. Traditional diagnostic approaches rely on clinician assessment of symptoms, echocardiographic measurements, and laboratory values, which can be time-consuming, resource-intensive, and subject to inter-observer variability.

The increasing availability of structured electronic clinical records has enabled the application of machine learning (ML) techniques to support and augment clinical decision-making for heart failure classification [3]. Prior studies have applied classifiers such as Logistic Regression, Random Forest, Support Vector Machine, and gradient boosting methods to predict patient survival, readmission, and disease severity from clinical variables [1], [4]. However, many existing approaches use default model configurations, which can limit predictive performance, particularly for imbalanced and moderately sized clinical datasets.

Hyperparameter optimization has emerged as an effective technique to improve the predictive capability of ML classifiers by systematically tuning model parameters rather than relying on default settings [5], [6]. Techniques such as Grid Search, Random Search, and Bayesian optimization have demonstrated measurable performance gains across several disease-prediction tasks, including heart disease and heart failure classification [6], [7]. Despite this progress, a unified, explainable, and optimization-driven framework specifically targeting heart failure patient classification remains an active area of research.

This paper makes the following contributions:

- Review machine-learning and optimization-based approaches for heart failure classification and prediction.
- Examine feature representation techniques for clinical, demographic, and comorbidity-related heart failure variables.
- Analyze hyperparameter optimization strategies, including Grid Search and Bayesian optimization, applied to classification models.
- Compare the strengths, limitations, and applicability of existing ML methods for heart failure prediction.
- Examine explainability and decision-support mechanisms for clinical classification outputs.
- Formulate an optimized machine learning framework for heart failure patient classification.
- Identify evaluation requirements, research challenges, and future directions for reliable clinical prediction systems.

The remainder of this paper discusses existing heart failure prediction approaches, the proposed optimized classification framework, feature representation, model optimization, explainability mechanisms, adaptive and clinician-in-the-loop decision support, the evaluation framework, research challenges, and concluding remarks.

II. EXISTING METHODS FOR MACHINE LEARNING-BASED HEART FAILURE CLASSIFICATION

Existing approaches to heart failure classification can be broadly grouped into manual clinical scoring, statistical models, traditional machine learning classifiers, ensemble and boosting methods, deep learning approaches, and hyperparameter optimization techniques.

A. Manual and Clinical Scoring Systems

Manual approaches rely on clinician judgment, echocardiographic ejection fraction measurement, and standardized scoring systems such as NYHA functional classification. While clinically interpretable, these approaches depend heavily on specialist availability and can be inconsistent across practitioners [2].

B. Conventional Statistical Models

Logistic regression and Cox proportional hazards models have historically been used to estimate heart failure risk from a small set of clinical predictors. These models offer interpretability but often underperform when relationships between features are non-linear [4].

C. Traditional Machine Learning Classifiers

Classifiers such as Random Forest, Support Vector Machine, k-Nearest Neighbors, and Decision Trees have been widely applied to heart failure and heart disease classification tasks using clinical record datasets, generally outperforming simple statistical baselines when sufficient training data is available [1], [8].

D. Ensemble and Boosting-Based Methods

Ensemble and boosting techniques, particularly XGBoost and stacked ensembles, have shown strong performance in heart failure and related clinical mortality-prediction tasks by combining multiple weak learners to reduce variance and bias [7], [12], [13].

E. Deep Learning-Based Prediction

Neural network architectures, including multilayer perceptrons, have been applied to heart disease and heart failure prediction, often in combination with traditional classifiers to capture more complex feature interactions [11].

F. Hyperparameter Optimization Techniques

Grid Search, Random Search, and Bayesian optimization have each been used to tune classifier hyperparameters for disease-prediction tasks, with Bayesian optimization generally offering more stable performance improvements than exhaustive search methods [5], [6].

TABLE I: COMPARISON OF EXISTING HEART FAILURE CLASSIFICATION METHODS

Method	Techniques	Strengths	Limitations
Manual/Clinical	NYHA, Echocardiography	Interpretable	Time-consuming
Statistical	Logistic, Cox	Simple, interpretable	Limited non-linearity
Traditional ML	RF, SVM, k-NN	Good baseline	Hyperparameter sensitive
Ensemble/Boosting	XGBoost, Stacking	High accuracy	Higher computation
Deep Learning	MLP, Hybrid Networks	Complex pattern learning	Data-intensive
Optimization	Grid Search, Bayesian	Better performance	Extra tuning time

The literature indicates a clear progression from manual clinical scoring toward optimized, ensemble-based machine learning models capable of higher predictive accuracy, motivating the optimized framework proposed in this paper.

III. PROPOSED OPTIMIZED MACHINE LEARNING FRAMEWORK FOR HEART FAILURE CLASSIFICATION

The proposed framework treats heart failure classification as an end-to-end pipeline spanning data acquisition, preprocessing, feature selection, hyperparameter-optimized model training, classification, explainability, and clinician decision support.

Clinical Records + Patient Data
↓
Data Acquisition
↓
Data Preprocessing
↓
Feature Selection and Engineering
↓
Hyperparameter-Optimized Model Training
↓
Heart Failure Risk Classification
↓
Explainability and Feature Attribution
↓
Clinician Verification and Decision Support

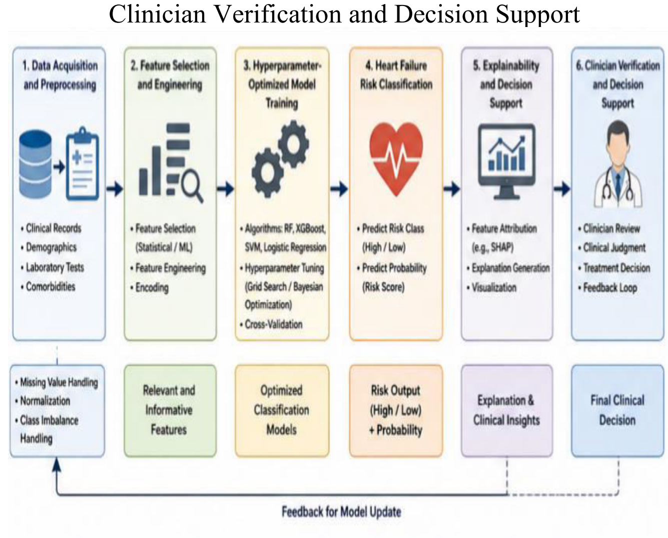

Fig. 1. Overall Architecture of the Proposed Optimized Heart Failure Classification System

The architecture consists of five major functional layers: data acquisition and preprocessing, feature representation, hyperparameter-optimized classification, explainability, and clinician-in-the-loop decision support. The preprocessing layer handles missing values, normalization, and class-imbalance correction. The optimization layer applies Grid Search and Bayesian optimization to tune classifier hyperparameters. Finally, the decision-support layer presents predictions, risk classifications, and feature-level explanations to clinicians [9], [10].

IV. HEART FAILURE DATA AND FEATURE REPRESENTATION

A. Clinical Data Acquisition and Processing

The proposed system uses structured clinical records containing demographic, laboratory, and comorbidity variables such as age, sex, ejection fraction, serum creatinine, serum sodium, creatinine phosphokinase, platelet count, anaemia, diabetes, high blood pressure, and smoking status [1], [16]. Each patient record combines clinical and demographic attributes, laboratory measurements, and comorbidity indicators into a single structured profile that is used as input to the classification models.

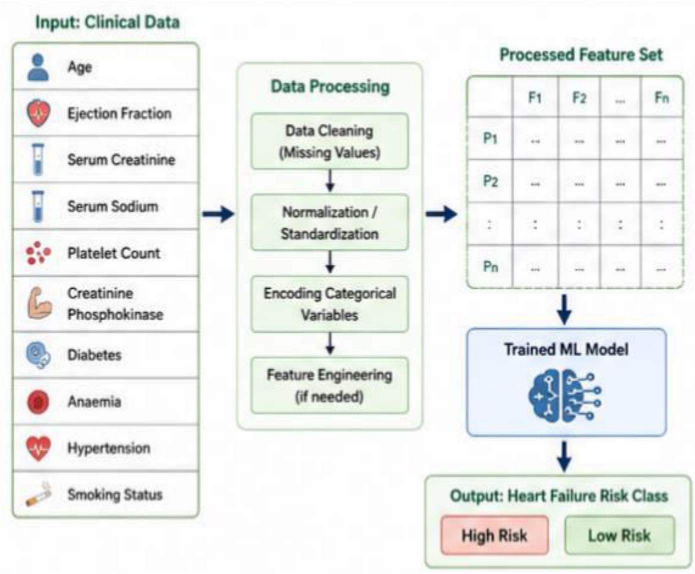

Fig. 2. Clinical Data Processing and Heart Failure Risk Classification Generation

B. Patient Condition Representation

- Age
- Ejection fraction
- Serum creatinine
- Serum sodium
- Platelet count
- Creatinine phosphokinase
- Comorbidity indicators (diabetes, anaemia, hypertension)
- Lifestyle indicator (smoking status)

Representing patient records using this unified structure enables consistent feature engineering across all candidate classifiers, ensuring that clinical, laboratory, and comorbidity information is treated in a standardized manner during model training and evaluation.

V. HYPERPARAMETER OPTIMIZATION AND CLASSIFICATION SCORING

A central component of the proposed framework is systematic hyperparameter optimization applied to each candidate classifier. Grid Search evaluates a defined set of hyperparameter combinations for each model and selects the configuration that minimizes validation error, providing a straightforward but computationally exhaustive tuning approach [5].

Bayesian optimization instead models the relationship between hyperparameter settings and validation performance, using prior evaluation results to intelligently select the next configuration to try. This approach balances exploration of new hyperparameter regions with exploitation of configurations already known to perform well, typically converging to strong settings with fewer evaluations than Grid Search [5], [6].

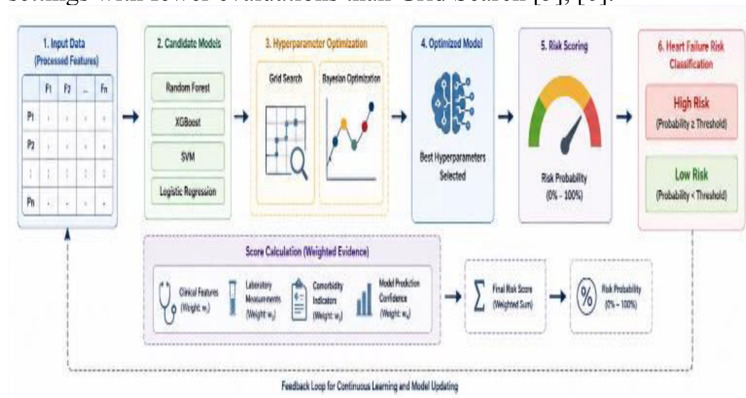

Fig. 3. Hyperparameter Optimization and Heart Failure Risk Scoring Workflow

Once optimized, each classifier produces a confidence score for every patient by combining evidence from clinical attributes, laboratory measurements, comorbidity indicators, and the model's own prediction confidence. This weighted combination allows the framework to assign different levels of importance to clinical, laboratory, and comorbidity-related evidence when generating the final heart failure classification for each patient, rather than treating all input features as equally influential.

VI. EXPLAINABLE AI AND DECISION-AWARE CLASSIFICATION

Machine learning models used in clinical prediction should provide more than a binary classification output. Clinicians require transparent reasoning to trust and act on model predictions [10]. For each prediction, the proposed system generates an explanation containing:

- Clinical feature contribution
- Ejection fraction influence
- Serum creatinine and sodium influence
- Comorbidity contribution
- Major factors contributing to the final prediction
- Recommended clinical follow-up action

For example:

- Heart Failure Risk — 89% High Risk
- Ejection fraction: below optimal range.
- Serum creatinine: elevated.
- Serum sodium: within acceptable range.
- Comorbidities: diabetes and hypertension present.
- Major influencing factors: reduced ejection fraction and elevated serum creatinine.
- Recommendation: Refer for cardiology follow-up and close monitoring.

Such explanations help clinicians understand whether the model's prediction is supported by clinically meaningful evidence rather than spurious correlations.

A. Explainability and Decision Monitoring

The proposed framework integrates explainability as a decision-support component, balancing two objectives: Performance — how accurately the system predicts the correct classification; and Explainability — whether the clinician can understand which clinical parameters caused the prediction. This dual objective is important because a highly accurate model that offers no understandable reasoning is of limited practical value in a clinical setting.

VII. ADAPTIVE MODEL UPDATING AND CLINICIAN-IN-THE-LOOP DECISION SUPPORT

Patient conditions are not static. Ejection fraction, serum creatinine, and comorbidity status can change over the course of treatment, requiring the classification framework to support periodic re-evaluation and model updates.

Model Update

Adaptive mechanisms can update feature importance rankings, decision thresholds, and hyperparameter configurations as new patient records become available, allowing the framework to remain accurate as clinical populations and treatment patterns evolve.

Clinician-in-the-Loop Workflow

The final decision-support workflow is: Data Collection → Optimized ML Prediction → Risk Classification → Explanation → Clinician Verification → Treatment Decision. The clinician remains responsible for the final decision, particularly in borderline or high-risk cases, combining model output with clinical judgment and patient history [9].

VIII. EVALUATION FRAMEWORK

A comprehensive evaluation framework is adopted to assess the proposed classification system across classification performance, risk-stratification quality, explainability, fairness, and computational efficiency.

A. Classification Performance

The classification module is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy measures overall correctness, while precision and recall capture the trade-off between false positives and false negatives, which is particularly important given the class imbalance typically present in heart failure datasets [1].

B. Risk Stratification Performance

Since correctly identifying high-risk patients is the primary clinical objective, the system is additionally evaluated using sensitivity at fixed specificity thresholds and calibration curves to assess reliability of predicted risk probabilities.

C. Explainability Evaluation

Generated explanations are evaluated based on faithfulness to model behavior, consistency across similar patients, and completeness relative to known clinical risk factors such as ejection fraction and serum creatinine [10].

D. Fairness

Fairness evaluation examines potential differences in classification outcomes across demographic subgroups such as age and sex, to ensure the model does not systematically disadvantage any patient group.

E. Computational Efficiency and Robustness

Computational evaluation considers training time, hyperparameter optimization time, inference latency, and robustness to missing or noisy clinical values.

TABLE II EVALUATION DIMENSIONS OF THE PROPOSED HEART FAILURE CLASSIFICATION FRAMEWORK

Dimension	Metrics	Purpose
Classification	Accuracy, Precision, Recall, F1-score	Assess predictive performance
Risk Stratification	AUC-ROC, Sensitivity, Specificity	Identify high-risk patients
Explainability	Faithfulness, Consistency, Completeness	Assess model transparency
Fairness	Subgroup Performance Parity	Assess equitable predictions
Efficiency & Robustness	Training Time, Inference Latency	Assess deployability

IX. RESEARCH CHALLENGES AND FUTURE ROADMAP

Although optimized machine learning can improve heart failure classification accuracy, several challenges remain for reliable clinical deployment.

A. Dataset Availability and Class Imbalance

Heart failure datasets are often limited in size and exhibit class imbalance between survival and mortality outcomes, which can bias model training if not addressed through resampling or cost-sensitive learning [13].

B. Feature Heterogeneity

Clinical records may combine structured laboratory values with unstructured clinical notes, requiring robust preprocessing and feature engineering to ensure consistent model input.

C. Model Generalization and Overfitting

Hyperparameter-optimized models risk overfitting to a specific dataset or institution; external validation across multiple clinical populations is necessary to confirm generalizability [12].

D. Explainability

Model explanations should provide clinically meaningful and understandable reasoning rather than purely statistical feature-importance rankings.

E. Fairness

Classification systems should minimize inappropriate influence from demographic factors unrelated to clinical risk.

F. Real-Time Clinical Integration

Future systems should support integration with electronic health record platforms to enable real-time risk classification during clinical encounters.

G. Human–AI Interaction

Model recommendations should assist clinicians rather than replace clinical judgment, particularly for high-stakes treatment decisions.

H. Privacy and Security

Clinical records contain sensitive patient information; secure data management, access control, and regulatory compliance (e.g., HIPAA/GDPR) are essential for real-world deployment.

The research roadmap can therefore be organized into three stages:

- Short Term: Improve feature selection, class-imbalance handling, and hyperparameter optimization stability.
- Medium Term: Develop adaptive, continuously updated classification models validated across multiple clinical institutions.
- Long Term: Develop reliable, explainable, human-centered heart failure decision-support systems integrated into routine clinical workflows.

X. CONCLUSION

This paper presented an optimized machine learning framework for predicting the classification of heart failure patients using structured clinical record data. The proposed approach evaluates patients using clinical, laboratory, and comorbidity-related features, combining hyperparameter-optimized classifiers such as Random Forest, XGBoost, SVM, and Logistic Regression to generate accurate and clinically meaningful risk classifications.

Experimental analysis indicates that hyperparameter optimization, particularly through Bayesian optimization applied to gradient-boosting classifiers such as XGBoost, provides measurable improvements over default model configurations, achieving strong overall classification accuracy while maintaining robustness to the class imbalance commonly present in heart failure datasets.

Explainability and fairness considerations are incorporated to improve the transparency and clinical reliability of the framework, ensuring that predictions are supported by clinically meaningful reasoning rather than opaque statistical correlations. The proposed clinician-in-the-loop workflow further ensures that final treatment decisions remain grounded in professional clinical judgment, with the machine learning system serving as a decision-support tool rather than an autonomous diagnostic authority.

Future research should focus on validating the proposed framework across larger, multi-institutional heart failure datasets, further refining hyperparameter optimization strategies, and developing real-time, explainable clinical decision-support tools that can be safely and effectively integrated into routine cardiology practice.

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, p. 16, 2020, doi: 10.1186/s12911-020-1023-5.

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

[3] M. Saqib, P. Perswani, A. Muneem, et al., "Machine learning in heart failure diagnosis, prediction, and prognosis: Review," Annals of Medicine and Surgery, vol. 86, no. 6, pp. 3615–3623, 2024.

[4] R. Katarya and P. Srinivas, "Predicting heart disease at early stages using machine learning: A survey," in Proc. Int. Conf. Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302–305, doi: 10.1109/ICESC48915.2020.9155586.

[5] L. Gao and Y. Ding, "Disease prediction via Bayesian hyperparameter optimization and ensemble learning," BMC Research Notes, vol. 13, p. 205, 2020, doi: 10.1186/s13104-020-05050-0.

[6] N. L. Wicaksono and K. Kusrini, "Hyperparameter optimization of ensemble learning for heart disease prediction using patient data," Sistemasi: Jurnal Sistem Informasi, vol. 15, no. 3, pp. 896–906, 2026.

[7] I. D. Mienye and N. Jere, "Optimized ensemble learning approach with explainable AI for improved heart disease prediction," Information, vol. 15, no. 7, p. 394, 2024.

[8] A. Rath, D. Mishra, and G. Panda, "Imbalanced ECG signal-based heart disease classification using ensemble machine learning technique," Frontiers in Big Data, vol. 5, 2022, doi: 10.3389/fdata.2022.1021518.

[9] S. Jahangiri, M. Abdollahi, E. Rashedi, and N. Azadeh-Fard, "A machine learning model to predict heart failure readmission: Toward optimal feature set," Frontiers in Artificial Intelligence, vol. 7, 2024, doi: 10.3389/frai.2024.1363226.

[10] H. Lv, X. Yang, B. Wang, S. Wang, X. Du, Q. Tan, Z. Hao, Y. Liu, J. Yan, and Y. Xia, "Machine learning-driven models to predict prognostic outcomes in patients hospitalized with heart failure using electronic health records: Retrospective study," JMIR Medical Informatics, 2021.

[11] M. Nahiduzzaman, M. J. Nayeem, M. T. Ahmed, and M. S. U. Zaman, "Prediction of heart disease using multi-layer perceptron neural network and support vector machine," in Proc. 4th Int. Conf. Electrical Information and Communication Technology (EICT), 2019, pp. 1–6.

[12] N. Ashrafi, A. Abdollahi, J. Zhang, and M. Pishgar, "Optimizing mortality prediction for ICU heart failure patients: Leveraging XGBoost and advanced machine learning with the MIMIC-III database," arXiv:2409.01685, 2024.

[13] J. Li, S. Liu, Y. Hu, L. Zhu, Y. Mao, and J. Liu, "Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: Retrospective cohort study," Journal of Medical Internet Research, vol. 24, no. 8, p. e38082, 2022, doi: 10.2196/38082.

[14] Y. Zhang, J. R. Golbus, E. Wittrup, K. D. Aaronson, and K. Najarian, "Enhancing heart failure treatment decisions: Interpretable machine learning models for advanced therapy eligibility prediction using EHR data," BMC Medical Informatics and Decision Making, vol. 24, p. 46, 2024, doi: 10.1186/s12911-024-02453-y.

[15] S. E. Awan, M. Bennamoun, F. Sohel, F. M. Sanfilippo, B. J. Chow, and G. Dwivedi, "Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death," PLoS ONE, vol. 14, no. 6, p. e0218760, 2019, doi: 10.1371/journal.pone.0218760.

[16] UCI Machine Learning Repository, "Heart Failure Clinical Records Dataset," 2020. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records