

## Predictive Model for Heart Failure Disease Using Ensemble Approach

Maishanu Abdallah<sup>1\*</sup>, Isah Bashar Bawa<sup>2</sup>, Tomori Rasheed Adekola<sup>3</sup>,  
Olukiran Oyeniike<sup>4</sup>, Samuel Ruth Medinat<sup>5</sup>

<sup>1,2</sup>Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria  
Department of Computer Science, University of Ilorin, Ilorin, Nigeria

<sup>4</sup>Lomond Lane, Kitchener, Ontario, Canada

<sup>5</sup>Department of Computer Science, Kogi State Polytechnic, Lokoja, Kogi State, Nigeria

\*Corresponding Author: [abdallahmaishanu@gmail.com](mailto:abdallahmaishanu@gmail.com)

**ABSTRACT:** Heart failure remains one of the leading causes of morbidity and mortality worldwide, necessitating the development of accurate and efficient predictive systems for early diagnosis and prevention. This study presents a Heart Failure Predictive Model using an Ensemble Learning Approach, designed to enhance diagnostic accuracy and reliability through the combination of multiple machine learning algorithms. Clinical datasets were obtained from the UCI Machine Learning Repository and Kaggle, containing key patient attributes such as age, blood pressure, ejection fraction, and serum creatinine levels. Data preprocessing involved cleaning, normalization, and the use of the SMOTEENN technique to correct class imbalance. Relevant features were selected using the Lion Optimization Algorithm (LOA) to improve model performance and reduce computational complexity. Five base classifiers, Artificial Neural Network (ANN), Decision Tree (C4.5), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM), were individually trained and later combined using ensemble techniques such as Voting. Experimental results demonstrated that ensemble methods significantly outperformed individual classifiers, achieving up to 99.75% accuracy using the boosting ensemble model. Performance metrics, including accuracy, precision, recall, F1-score, and kappa statistic, confirmed the robustness and reliability of the model. The study concluded that ensemble learning provides a powerful, data-driven approach for early heart failure prediction, enabling timely clinical interventions and improved patient outcomes. The developed model serves as a foundation for integrating intelligent diagnostic systems into healthcare environments, contributing to the advancement of predictive medicine and the practical application of artificial intelligence in medical diagnostics.

**KEYWORDS:** Ensemble Model, Heart Failure Disease, Mortality, Predictive Model

### I. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death globally, accounting for approximately 17.9 million deaths annually, which represents about 32% of all global deaths (World Health Organization (The & Killer, 2023). Among these, heart failure (HF) is a major public health concern that contributes significantly to morbidity, mortality, and healthcare costs (Shahim et al., 2023). Heart failure, often described as a clinical syndrome characterized by the heart's inability to pump blood efficiently to meet the body's metabolic demands (Gitif et al., 2021). It is a progressive condition that, if not detected early, leads to frequent hospitalizations and reduced quality of life. Early detection of heart failure is vital for timely intervention and improved patient outcomes (Zeru, 2018). However, predicting heart failure can be challenging due to its complex etiology and overlap with other cardiovascular and systemic conditions (Id et al., 2024; Addissouky et al., 2024). Traditional methods of prediction typically rely on a combination of clinical examination, echocardiographic findings, and laboratory biomarkers such as B-type natriuretic

peptide (BNP). While effective, these methods can be costly, time-consuming, and not always accessible in resource-limited settings (Lala et al., 2025).

In recent years, the integration of artificial intelligence (AI) and machine learning (ML) techniques into medical prediction has opened new avenues for improving clinical decision-making (Obiany et al., 2025). These computational techniques enable the analysis of large, multidimensional datasets to uncover hidden patterns and predictive relationships that may not be apparent through traditional statistical approaches (Theodorakopoulos & Theodoropoulou, 2024). In the context of heart failure, machine learning models have been employed to predict disease onset, readmission risk, and mortality by analyzing structured and unstructured patient data. Despite the promising results of machine learning models, they still suffer from limitations related to overfitting, sensitivity to noise, and limited generalizability, especially when applied to heterogeneous clinical data (Tran et al., 2025). This has led to a growing interest in ensemble learning approaches, which combine multiple base learners to produce a more robust and accurate predictive model. Ensemble methods such as boosting and stacking have shown superior performance in various healthcare applications by reducing variance, bias, or both. Ensemble classification models leverage the strengths of individual learners while mitigating their weaknesses (Zhu, 2024). For instance, Random Forest aggregates multiple decision trees trained on different data subsets, which enhances stability and reduces the risk of overfitting. Boosting techniques, on the other hand, sequentially train models by focusing on previously misclassified instances, thereby improving classification accuracy. These ensemble approaches have shown significant potential in diagnosing and predicting cardiovascular conditions, including heart failure, with improved accuracy and interpretability (Dwi et al., 2024; Scifoni et al., 2025).

Recent studies have explored the use of ensemble models in early heart failure detection (Gour et al., 2025). For example, the development of a predictive model using a combination of Random Forest and gradient boosting techniques, achieving high accuracy in identifying patients at risk of HF hospitalization. The application of feature selection is significant in the prediction of heart failure due to its reduction in model complexity, speeding up training, preventing overfitting, and boosting accuracy by selecting only the most relevant features, improving interpretability and efficiency in fields like IoT, healthcare, and bioinformatics (Hussain et al., 2024). Among several nature-inspired optimization algorithms, such as Bat Optimization Algorithm (BOA), Ant Colony Optimization (ACO), Genetic Algorithm (GA), Lion Algorithm, and Gravitational Search Algorithm (GSA) for feature selection techniques (Alsamia et al., 2024; Hemdanou et al., 2024), the LOA has been identified to be fast in convergence, better memory efficiency, effective global optimization, robustness, and high accuracy in complex engineering and AI tasks, leveraging behaviors like hunting, territorial defense, and pride dynamics for effective exploration and exploitation (Improved et al., 2018). The paper presented a predictive model for heart failure by applying LOA and introducing an ensemble model through voting to enhance the prediction of heart failure.

## II. RELATED WORKS

Singh and Singh (2025) explored diverse Heart Disease predictions based on deep learning, machine learning, and optimization methods. Recently, many researchers have used deep learning and machine learning algorithms to assist healthcare professionals in heart disease diagnoses. The study also examined several optimization algorithms and their performance. Ultimately, the review suggested that optimization-based HDP methods can play a crucial role in helping doctors predict heart disease early and recommend appropriate treatments. Shourya et al. (2025) presented work using various techniques of machine learning and data mining to analyze complex medical datasets. The work specifically considered supervised learning algorithms, including Naïve Bayes, decision trees, KNN, and random forests. The employed data set comes from the Cleveland database in the UCI repository, which deals with data from patients affected by heart diseases. Although the dataset has 303 instances and 76 attributes, the analysis is performed only on the selected 14 key attributes to estimate the performance of a few models. Udhan and Patil (2025) reviewed existing work that applied Deep Learning (DL) and ML techniques to predict heart disease. From the study, it has been discovered that most of the work has used supervised ML algorithms, which include Support Vector Machines (SVM), Gradient Boosting Classifiers (GB), Decision Trees (DT), Random Forest (RF),

and Logistic Regression (LR), have been employed on the UCI Machine Learning Repository (Heart) dataset to predict cardiac conditions. The accuracy of these algorithms varies, with studies reporting success rates between 88% and 95%. This review explores the factors influencing these outcomes, contributing to a better understanding of ML-based heart disease prediction models.

El-Sofany et al (2024) used different feature selection approaches to develop an accurate ML algorithm for early heart disease prediction. Chi-square, ANOVA, and mutual information methods were collected. The three feature groups chosen were SF-1, SF-2, and SF-3. The study employed ten learning algorithms: support vector machines (SVM), XGBoost, bagging, decision trees (DT), and random forests (RF), to determine the most accurate technique and feature subset fit. Evaluation of the proposed heart disease prediction technique using a private dataset, a public dataset, and different cross-validation methods. The Synthetic Minority Oversampling Technique (SMOTE) removes inconsistent data and discovers the machine learning algorithm that achieves the most accurate heart disease predictions. The work employed the most effective ML algorithm to develop a mobile app that instantly predicts heart disease based on the input symptoms. The experimental results showed that the XGBoost algorithm performed optimally when applied to the combined datasets and the SF-2 feature subset. It had 97.57% accuracy, 96.61% sensitivity, 90.48% specificity, 95.00% precision, a 92.68% F1 score, and a 98% AUC. Ayyavvu et al. (2023) evaluated the accuracy of machine learning algorithms, specifically nearest neighbor, decision tree, linear regression, and support vector machine (SVM), in predicting heart disease. The UCI repository dataset was used for training and testing, and employed the Python programming language through the Anaconda Jupyter notebook for implementation. The experimental results indicated an accuracy rate of 95.7% using the heart disease prediction model, leveraging its extensive libraries and header files to ensure accuracy and precision in our analysis. Shiwangi et al. (2023) proposed a model to apply to a real-world dataset of 70,000 instances from Kaggle. Models were trained on data that were divided into 80% and 20 and achieved accuracy as follows: decision tree: 86.37% (with cross-validation) and 86.53% (without cross-validation), XGBoost: 86.87% (with cross-validation) and 87.02% (without cross-validation), random forest: 87.05% (with cross-validation) and 86.92% (without cross-validation), multilayer perceptron: 87.28% (with cross-validation) and 86.94% (without cross-validation). The proposed models have AUC (area under the curve) values: decision tree: 0.94, XGBoost: 0.95, random forest: 0.95, and multilayer perceptron: 0.95. Finally, the multilayer perceptron with cross-validation outperformed all other algorithms with an accuracy of 87.28. Mohi Uddin et al. (2023) developed a model to predict the chances of heart disease and classify patients' risk level by applying different learning algorithm techniques; Decision Tree (DT), Ada-Boost Classifier (AB), Extra trees Classifier (ET), Support vector Machine (SVM), Gradient boost, MLP, extreme gradient boost (XGB), Random Forest (RF), KNN, and LR. Three different datasets were combined to train and test the proposed system. The experimental results revealed that, when compared to other ML algorithms, the Decision Tree algorithm performed better than existing models with an accuracy value of 99.16%.

### III. METHODOLOGY

The heart failure predictive model will employ stepwise approaches. Two datasets, the UCI Machine learning repository and Kaggle online datasets, will be used. The preprocessing will be performed on the acquired datasets. The preprocessed attributes were passed to the optimization algorithm for attribute selection via the Lion Optimization Algorithm. The optimized attributes and the preprocessed attributes were passed to some selected learning techniques, such as ANN, C4.5, LR, KNN, and SVM, with an ensemble method through voting for robust classification. Finally, the results were evaluated in terms of accuracy, precision, recall, F1-Score, and Time taken to build the model. The framework of the heart failure prediction model is shown in Figure 1.

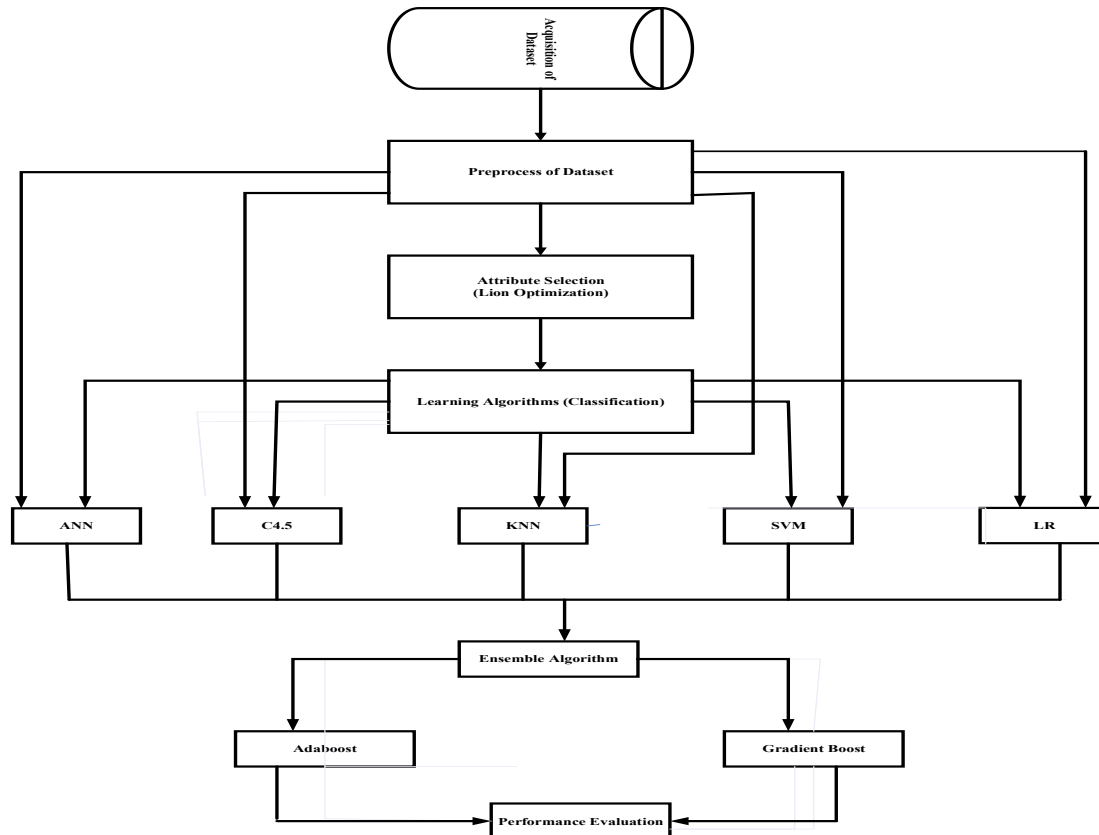


Figure 1: Heart Failure Predictive Model

**A. Acquisition and Description of Heart Failure Datasets**

The study used online available datasets; the UCI Machine learning repository and the Kaggle Dataset to evaluate the performance of the developed heart failure predictive model. The UCI machine learning dataset contains the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 13 clinical features and 299 instances. Also, the heart failure clinical report dataset was obtained from Davide Chicco, Giuseppe Jurman: “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making. A sample of the attributes is given in Table 1

Table 1: Sample of Attributes for Heart Failure Dataset

75,0,582,0,20,1,265000,1.9,130,1,0,4,1
55,0,7861,0,38,0,263358.03,1.1,136,1,0,6,1
65,0,146,0,20,0,162000,1.3,129,1,1,7,1
50,1,111,0,20,0,210000,1.9,137,1,0,7,1
65,1,160,1,20,0,327000,2.7,116,0,0,8,1
90,1,47,0,40,1,204000,2.1,132,1,1,8,1
75,1,246,0,15,0,127000,1.2,137,1,0,10,1
60,1,315,1,60,0,454000,1.1,131,1,1,10,1
65,0,157,0,65,0,263358.03,1.5,138,0,0,10,1
80,1,123,0,35,1,388000,9.4,133,1,1,10,1

75,1,81,0,38,1,368000,4,131,1,1,10,1
62,0,231,0,25,1,253000,0.9,140,1,1,10,1
45,1,981,0,30,0,136000,1.1,137,1,0,11,1
50,1,168,0,38,1,276000,1.1,137,1,0,11,1
49,1,80,0,30,1,427000,1,138,0,0,12,0
82,1,379,0,50,0,47000,1.3,136,1,0,13,1
87,1,149,0,38,0,262000,0.9,140,1,0,14,1
45,0,582,0,14,0,166000,0.8,127,1,0,14,1
70,1,125,0,25,1,237000,1,140,0,0,15,1
48,1,582,1,55,0,87000,1.9,121,0,0,15,1
65,1,52,0,25,1,276000,1.3,137,0,0,16,0

### B. Data Preprocessing

Data preprocessing is an important phase in building a reliable and high-performing predictive model. It ensures that the dataset is clean, consistent, and well-formatted for training machine learning algorithms. The preprocessing step applied to the Heart Failure Clinical Record Dataset is given in Figure 3.2.

### C. Feature Selection Using LOA

The procedure for the selection of features for the Heart Failure Clinical Dataset is given as follows:

Step 1: Start

Step 2: Initialization

- Create an initial population of *N* lions (candidate solutions).
- Randomly assign positions in the search space (within variable bounds).
- Assign genders to lions (male/female) according to a chosen male ratio (e.g., 20–30%).
- Partition the population into several *prides* and a set of *nomads* (usually by percentage or clustering).

Step 3: Evaluate fitness

- Compute the objective value (fitness) for every lion. Keep track of the best solution(s) globally and per pride.

Step 4: Pride operations Hunting (female-driven local search)

- In each pride, selected females perform cooperative hunting to generate new candidate positions.
- Hunting typically builds new solutions by combining the positions of several best females and sometimes the pride male(s). This is the main exploitation operator (local search). If hunting produces better solutions, replace weaker pride members.

Step 5: Roaming (male exploration)

- Resident males (and sometimes females) perform roaming: stochastic moves that explore nearby or distant regions (e.g., random perturbations scaled by a parameter). Roaming increases diversity and helps escape local optima.

Step 6: Mating (reproduction/crossover)

- Mating between resident male(s) and female(s) in a pride (and sometimes nomad × pride) creates offspring (new candidate solutions).
- Mating uses crossover-like operators to combine parents. Offspring replace the worst lions if they are fitter. Fertility and probability of mating are parameters.

Step 7: Nomad Behaviour & Takeovers

- *Nomads* roam widely (high exploration). They can attack a pride to try to take it over: if a nomad male is fitter than some pride male(s), it may replace them (simulating territorial takeover). This brings strong new solutions into pride and removes poor performers.

Step 8: Migration between prides

- Some lions may migrate between prides (or from nomads to prides) according to a small migration probability, which helps share information across groups and avoid premature convergence.

Step 9: Cubs' survival (replacement policy)

- Offspring may have a survival probability; weak cubs are removed. This mimics selection pressure. Replacement rules (which parents are replaced) vary by implementation.

Step 10: Update bests and enforce population limits

- After the above operators, re-evaluate fitnesses, update global/pride bests, and enforce fixed population size (remove worst if over).

Step 11: Stopping condition

- Repeat steps 3–9 until stopping criteria: max iterations, CPU time, or no improvement for a number of iterations.

Step 12: Stop

## 4. RESULTS AND DISCUSSION

### A. Experimental Analysis

The heart failure prediction model was developed and implemented in the Python programming environment using Google Colab, as shown in Figure 2. The Colab command window serves as the console interface where outputs from data preprocessing, feature selection, model training, and evaluation are displayed.

### B. Loading of the Dataset

The loading of the Heart Failure Clinical Records Dataset was carried out, which the dataset contains medical data from 5,000 patients who had experienced heart failure, as shown in Figure 3.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	anaemia	creatinine	diabetes	ejection_t	high_blo	platelets	serum_cr	serum_so	sex	smoking	time	DEATH_EVENT	
55	0	748	0	45	0	263358	1.3	137	1	1	88	0	
65	0	56	0	25	0	305000	5	130	1	0	207	0	
45	0	582	1	38	0	319000	0.9	140	0	0	244	0	
60	1	754	1	40	1	328000	1.2	126	1	0	90	0	
95	1	582	0	30	0	461000	2	132	1	0	50	1	
70	0	232	1	30	0	302000	1.2	132	1	0	210	0	
63	1	122	1	60	0	172000	1.2	145	0	0	147	0	
70	1	171	0	50	1	358000	0.9	141	0	0	196	0	
50	0	482	1	30	0	300000	0.9	132	1	0	109	0	
53	1	446	0	45	1	274000	1	133	1	0	215	0	
60.667	1	151	1	40	1	257000	1	136	0	0	172	0	
72	1	110	0	25	0	237000	1	140	0	0	65	1	
70	0	2695	0	40	0	277000	1	137	1	1	247	0	
64	0	1610	0	60	0	263358	1	137	1	1	113	0	
65	0	335	0	35	1	235000	0.8	136	0	0	120	0	
60	1	34	0	20	0	204000	0.7	139	1	1	73	1	
75	0	99	1	20	1	70000	2.3	134	1	0	162	1	
66	1	68	1	38	1	303000	1	136	1	0	95	0	
72	1	110	0	25	0	237000	1	140	0	0	65	1	
60	1	47	1	20	0	336000	0.7	139	1	0	73	1	
63	1	103	1	35	0	287000	0.9	136	1	1	270	0	
58	1	200	0	40	0	164000	1	139	0	0	91	0	
30	1	111	0	20	0	264000	1.9	137	1	1	7	1	
70	1	67	0	35	0	223000	2.7	138	1	1	54	0	
42	1	86	0	35	0	237000	1.1	139	0	0	201	0	
53	1	1807	0	38	0	330000	1.4	137	1	1	209	0	

Figure 3: Heart Failure Clinical Record Dataset

### C. Preprocessing of the Dataset

Before training the predictive model, the dataset was preprocessed to ensure consistency and improve model performance. The target variable, DEATH\_EVENT, which indicates patient survival status, was separated from the feature set. All remaining columns representing clinical attributes were retained as input features. To bring all features to a common scale and eliminate the effect of differing units, standardization was applied using the StandardScaler function from sklearn. Preprocessing module as discussed in Figure 4.

```
Preprocessing and features Normalization

X = df.drop("DEATH_EVENT", axis=1)
y = df["DEATH_EVENT"]

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Figure 4: Data Preprocessing

#### D. Balance Classes using SMOTEENN & Train/Test split

Class imbalance was addressed in the heart failure dataset using SMOTEENN, a hybrid resampling technique that combined SMOTE (Synthetic Minority Oversampling Technique) with Edited Nearest Neighbors (ENN). This approach improved classifier performance by oversampling minority instances and cleaning noisy samples from the majority class. Figure 5 shows the Python implementation of this stage:

```
Balance classes using SMOTEENN & Train/Test split

balancer = SMOTEENN(random_state=42)
X_bal, y_bal = balancer.fit_resample(X_scaled, y)

X_train, X_test, y_train, y_test = train_test_split(X_bal, y_bal, test_size=0.2, random_state=42)
```

Figure 5: Balance classes using SMOTEENN

#### E. Ensemble Approach of Classifiers

The ensemble classification algorithms were employed to model heart failure prediction, each representing a different learning strategy. The learning algorithms were improved by introducing a voting approach of an ensemble model. The Python implementation of the ensemble classification models is shown in Figure 6.

```

models = {
    "ANN": best_ann,
    "C4.5": DecisionTreeClassifier(criterion="entropy", max_depth=5),
    "LR": LogisticRegression(max_iter=500, solver='liblinear'),
    "KNN": KNeighborsClassifier(n_neighbors=5),
    "SVM": SVC(kernel="rbf", C=1, gamma='scale', probability=True),
    "XGBoost": XGBClassifier(eval_metric='logloss', n_estimators=100)
}

```

Figure 6: Ensemble Classification Models

A soft voting ensemble was constructed by combining six individual classifiers: Logistic Regression, Decision Tree (C4.5), Support Vector Machine, K-Nearest Neighbors, Artificial Neural Network, and XGBoost. The VotingClassifier was configured with voting='soft', allowing it to average the predicted class probabilities from each base estimator, which typically enhances performance by leveraging complementary strengths of the models. The ensemble was trained on the balanced training set. The Python implementation of the soft voting ensemble method is shown in Figure 7.

```

Voting Ensemble
ensemble = VotingClassifier(
    estimators=[
        ('lr', LogisticRegression(max_iter=500, solver='liblinear')),
        ('dt', DecisionTreeClassifier(criterion="entropy", max_depth=5)),
        ('svm', SVC(kernel="rbf", C=1, gamma="scale", probability=True)),
        ('knn', KNeighborsClassifier(n_neighbors=5)),
        ('ann', best_ann),
        ('xgb', XGBClassifier(eval_metric="logloss", n_estimators=100))
    ],
    voting="soft"
)

start = time.time()
ensemble.fit(X_train, y_train)
ensemble_preds = ensemble.predict(X_test)
end = time.time()

results["Voting_ensemble"] = {
    "Accuracy": accuracy_score(y_test, ensemble_preds),
    "Precision": precision_score(y_test, ensemble_preds),
    "Recall": recall_score(y_test, ensemble_preds),
    "F1 Score": f1_score(y_test, ensemble_preds),
    "Kappa": cohen_kappa_score(y_test, ensemble_preds),
    "Time taken": round(end - start, 3)
}

```

Figure 7: Voting Ensemble Method

### F. Result of Heart Failure Predictive Model

The performance evaluation of the individual classifiers and the developed voting ensemble model on the heart failure prediction task is clearly shown in this section. Each model was assessed using evaluation metrics: accuracy, precision, recall, F1-score, and kappa, along with the time taken for training and prediction.

### G. Result of Base Classifier (without feature selection/no ensemble)

The result of the base classifier is shown in Table 2, which indicates that when the ensemble approach was not applied with boosting to the heart failure classification phase.

Table 2: Result of Base Classifier (without feature selection / no ensemble method)

Model	Accuracy(%)	Precision	Recall	F1 Score	Kappa -statistic	Time Taken(S)
ANN	97.33	0.9734	0.9761	0.9931	0.9865	3.777
C4.5	87.58	0.8758	0.8705	0.8995	0.7918	0.013
LR	77.28	0.7720	0.7960	0.7919	0.7862	0.011
KNN	97.48	0.9704	0.9701	0.9947	0.9895	0.115

SVM	96.33	0.9565	0.9601	0.9633	0.9611	2.116
-----	-------	--------	--------	--------	--------	-------

From Table 2, the lowest accuracy of 77.28% was obtained in LR, the highest accuracy of 97.48% was recorded in KNN, the lowest precision of 0.7720 was obtained in LR, the highest precision of 0.9734 was recorded in ANN, the lowest recall of 0.7960 was obtained in LR, the highest recall of 0.9761 was recorded in ANN, the lowest F1-score of 0.7919 was obtained in LR, the highest F1-score was recorded in 0.9931, the lowest kappa-statistic of 0.7862 was obtained in LR, the highest kappa-statistic of 0.9895 was recorded in KNN, the lowest time of 0.011s was obtained in LR, the prolonged time of 3.777s was recorded in ANN. The high accuracy, precision, recall, and F1-scores demonstrated by the best-performing models (KNN and ANN) indicate strong reliability in correctly identifying true clinical cases, which is essential for early diagnosis and timely intervention. The high recall of ANN reduces the risk of missed cases, while KNN's high accuracy and kappa value show strong agreement with real clinical outcomes. These models can therefore support clinicians in improving diagnostic decisions, reducing errors, and enhancing patient care.

#### H. Result of Ensemble ( with feature selection / no ensemble method)

The result of the ensemble classifier is shown in Table 3, which indicates that when the feature selection was applied, but no ensemble approach was used in the heart failure classification phase.

**Table 3:** Ensemble (with feature selection / no voting)

Model	Accuracy (%)	Precision	Recall	F1 Score	Kappa -statistic	Time Taken(s)
ANN	99.32	0.9939	0.9924	0.9932	0.9865	2.70
C4.5	89.58	0.8591	0.9439	0.8995	0.7918	0.14
LR	79.31	0.7871	0.7967	0.7919	0.7884	0.8
KNN	99.47	0.9985	0.9909	0.9947	0.9895	0.11
SVM	96.33	0.9499	0.9772	0.9634	0.9266	1.12

From Table 3, the lowest accuracy of 79.31% was obtained in LR, the highest accuracy of 99.32% was recorded in KNN, the lowest precision of 0.7871 was obtained in LR, the highest precision of 9939 was recorded in ANN, the lowest recall of 0.7867 was obtained in LR, the highest recall of 0.9924 was recorded in ANN, the lowest F1-score of 0.7919 was obtained in LR, the highest F1-score of 0.9932 was recorded in ANN, the lowest kappa-statistic of 0.7884 was obtained in LR, the highest kappa-statistic of 0.9895 was recorded in KNN, the lowest time of 0.8s was obtained in LR, the prolonged time of 2.70s was recorded in ANN. The results show that **KNN** and **ANN** models provide highly accurate and reliable predictions, with ANN demonstrating the strongest precision, recall, and F1-score, and KNN achieving the highest accuracy and agreement. Their strong performance indicates they are suitable for clinical decision-support systems where minimizing diagnostic errors is critical. In contrast, Logistic regression shows the weakest performance across all metrics, making it less appropriate for clinical use due to a higher risk of misclassification. Although LR is faster computationally, the superior accuracy of KNN and ANN makes them more clinically valuable for improving early detection, reducing errors, and supporting better patient outcomes.

#### I. Result of Ensemble (feature selection/voting)

The result of the ensemble classifier is shown in Table 4, which indicates that when the feature selection was applied with a boosting ensemble approach to the heart failure classification phase was improved.

**Table 4:** Ensemble (with feature selection/voting)

Model	Accuracy (%)	Precision	Recall	F1 Score	Kappa -statistic	Time Taken(s)
ANN	99.75	0.9970	0.9910	0.9980	0.9810	47.0
C4.5	90.10	0.9012	0.9010	0.8894	0.7567	0.18

LR	81.35	0.8135	0.8301	0.8023	0.7512	0.22
KNN	99.47	0.9947	0.9947	0.9924	0.9702	0.25
SVM	96.33	0.9499	0.9772	0.9634	0.9266	2.12

From Table 4, the lowest accuracy of 81.35% was obtained in LR, the highest accuracy of 99.47% was recorded in KNN, the lowest precision of 0.8135 was obtained in LR, the highest precision of 0.9947 was recorded in KNN, the lowest recall of 0.83 was obtained in LR, the highest recall of 0.9947 was recorded in KNN, the lowest F1-score of 0.8023 was obtained in LR, the highest F1-score of 0.9932 was recorded in ANN, the lowest kappa-statistic of 0.7512 was obtained in LR, the highest kappa-statistic of 0.9810 was recorded in ANN, the lowest time of 0.18s was obtained in C4.5, the prolonged time of 47s was recorded in ANN.

## DISCUSSION OF RESULTS

The results of the predictive model for heart failure showed that KNN and ANN were the strongest individual classifiers, achieving accuracies of 97.48% and 97.33%, while Logistic Regression performed the worst across all metrics. Applying feature selection with LOA improved the performance of all models, with ANN and KNN again showing the highest precision, recall, F1-scores, and kappa values. The best performance was achieved when both feature selection and ensemble voting were applied: the ensemble-enhanced ANN reached 99.75% accuracy, outperforming all other models, while KNN remained highly effective at 99.47%. Although the ensemble ANN required more computation time, the combination of optimized features and ensemble learning significantly improved prediction reliability. Overall, the findings confirm that ensemble methods—especially after feature optimization—provide the most accurate and robust solution for heart failure prediction.

## IV. CONCLUSION

This study developed an effective heart failure predictive model using a combination of feature optimization and ensemble learning techniques. By integrating datasets from UCI, applying SMOTEENN for class balancing, and using the Lion Optimization Algorithm for feature selection, the model achieved substantial improvements in predictive accuracy. Among the evaluated algorithms, ANN and KNN consistently performed best, and their performance was further enhanced through ensemble voting. The ensemble-based ANN achieved the highest accuracy of 99.75%, demonstrating the superior capability of ensemble learning in reducing misclassification and improving clinical reliability. Overall, the results confirmed that combining optimized feature selection with ensemble methods provides a robust, accurate, and scalable predictive framework capable of supporting early diagnosis of heart failure and aiding clinical decision-making. Future research may focus on deploying the model in real-world clinical systems and extending it to broader patient populations.

## REFERENCES

- Addissouky, T. A., El, I., El, T., Ali, M. M. A., Hasen, M., Alubiady, S., & Wang, Y. (2024). Recent developments in the diagnosis, treatment, and management of cardiovascular diseases through artificial intelligence and other innovative approaches. *Journal of Biomed Research*, *1*, 29–40. <https://doi.org/10.46439/biomedres.5.041>
- Alsamia, S., Koch, E., Albedran, H., & Ray, R. (2024). Adaptive Exploration Artificial Bee Colony for Mathematical Optimization. *AI*, *5*, 2218–2236.
- Ayyavvu, K., Panneer, B. I., Sreenivasan, A., & Muthukrishnan, A. K. A. (2023). Heart disease prediction using machine learning. *AIP Conference Proceedings*, *2857*(1), 1–6. <https://doi.org/10.1063/5.0165188>
- Dwi, N., Nur, A., Tri, J., & Arai, K. (2024). Ensemble learning approaches for predicting heart failure outcomes : A comparative analysis of feedforward neural networks, random forest, and XGBoost.

- Applied Engineering and Technology*, 3(3), 173–184.
- El-Sofany, H., Bouallegue, B., & El-Latif, Y. M. A. (2024). A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific Reports*, 14(1), 23277. <https://doi.org/10.1038/s41598-024-74656-2>
- Gour, S., Randa, R., Pandagre, K. N., Neema, A., & Jat, S. (2025). A Comprehensive Review on Predicting Heart Failure with Ensemble Learning : Integrating Environmental Variables and Health Perspectives A Comprehensive Review on Predicting Heart Failure with Ensemble Learning : Integrating Environmental Variables and Hea. *International Journal of Environmental Sciences*, 11(4S), 959–973.
- Gtif, I., Bouzid, F., Charfeddine, S., Abid, L., & Kharrat, N. (2021). Heart failure disease : An African. *Archives of Cardiovascular Disease*, 114, 681–690. <https://doi.org/10.1016/j.acvd.2021.07.001>
- Hemdanou, A. L., Sefian, M. L., Ahtoun, Y., & Tahiri, I. (2024). Computers and Education : Artificial Intelligence Comparative analysis of feature selection and extraction methods for student performance prediction across different machine learning models. *Computers and Education: Artificial Intelligence*, 7(September), 100301. <https://doi.org/10.1016/j.caeai.2024.100301>
- Hussain, I., Qureshi, M., Ismail, M., Iftikhar, H., Zywio, J., & L, J. L. (2024). Heliyon Optimal features selection in the high-dimensional data based on a robust technique : Application to different health databases. *Heliyon*, 10(September).
- Id, I. D., Szubert, B., Murphy, C., Id, K. B., & Id, J. L. (2024). Early detection of heart failure using patient longitudinal electronic health records. *PLoS ONE*, 1–15. <https://doi.org/10.1371/journal.pone.0314145>
- Improved, A., Lion, A., & Algorithm, O. (2018). An Improved Ant Lion Optimization Algorithm and. *Energies*, 11, 1–15. <https://doi.org/10.3390/en11010095>
- Lala, A., Beavers, C., Blumer, V., Brewer, L., Oliveira-gomes, D. De, Dunbar, S., Every, H., Ferraro, R., Ky, B., Januzzi, J., Marvel, F., Mentz, R., Michos, E., Narula, J., Nasir, K., Natarajan, P., Ann, L., Rodriguez, F., Shapiro, M. D., ... Gulati, M. (2025). American Journal of Preventive Cardiology The continuum of prevention and heart failure in cardiovascular medicine : A joint scientific statement from the Heart Failure Society of America and the American Society for Preventive Cardiology. *American Journal of Preventive Cardiology*, 24, 1–24. <https://doi.org/10.1016/j.ajpc.2025.101069>
- Mohi Uddin, K. M., Ripa, R., Yeasmin, N., Biswas, N., & Dey, S. K. (2023). Machine learning-based approach to the diagnosis of cardiovascular disease using a combined dataset. *Intelligence-Based Medicine*, 7(April), 100100. <https://doi.org/10.1016/j.ibmed.2023.100100>
- Obianyo, C. M., Nzeyimana, P., Alehegn, H. Z., Drammeh, A., & Lawal, A. (2025). The Integration of Artificial Intelligence ( AI ) and Machine Learning ( ML ) in Diagnostics and Personalized Medicine. *Biomedical Journal of Scientific & Technical Research*, 61(5), 53977–53993. <https://doi.org/10.26717/BJSTR.2025.61.009650>
- Scifoni, E., Alshayegi, M. H., & Abed, S. (2025). Heart disease prediction by tabular modeling with deep learning network and interpretability Heart disease prediction by tabular modeling with deep learning network and interpretability. *IOP*, 6, 1–24.
- Shahim, B., Kapelios, C. J., & Savarese, G. (2023). Global Public Health Burden of Heart Failure : An Updated Review. *Clinical Syndromes*, 1–8.
- Shiwangi, K. M., Sandhu, J. K., & Sahu, R. (2023). Effective Heart-Disease Prediction by Using Hybrid Machine Learning Technique. *Proceedings of the International Conference on Circuit Power and Computing Technologies, ICCPCT 2023*, 1670–1675. <https://doi.org/10.1109/ICCPCT58313.2023.10245785>
- Shourya Singh<sup>1</sup>, Vishesh Singh<sup>2</sup>, Youg Khanna<sup>3</sup> <sup>1,2,3</sup>SCSE. (2025). Heart disease prediction using machine learning. *International Journal for Multidisciplinary Research (IJFMR)*, 7(2), 1–6. <https://doi.org/10.1063/5.0165188>
- Singh, N., & Singh, B. P. (2025). Review of Heart Disease Prediction using Machine Learning. *International Journal of Current Engineering and Technology*, 15(2), 165–169.

- The, C., & Killer, O. N. E. (2023). *World Heart Report 2023: Confronting the World Number*.
- Theodorakopoulos, L., & Theodoropoulou, A. (2024). A State-of-the-Art Review in Big Data Management Engineering : Real-Life Case Studies, Challenges, and Future Research Directions. *Eng*, 5, 1266–1297.
- Tran, A. T., Zeevi, T., & Payabvash, S. (2025). Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging. *BioMeddInformatics*, 5(20), 1–26.
- Udhan, S., & Patil, B. (2025). A systematic review of Machine learning techniques for Heart disease prediction. *International Journal of Next-Generation Computing*, 186(63), 229–239. <http://ijngc.perpetualinnovation.net/index.php/ijngc/article/view/208>
- Zeru, M. A. (2018). Assessment of major causes of heart failure and its pharmacologic management among patients at Felege Hiwot referral hospital in Bahir Dar, Ethiopia. *Journal of Public Health and Epidemiology*, 10, 326–331. <https://doi.org/10.5897/JPHE2018.1046>
- Zhu, Z. (2024). Systematic Optimization of Overfitting Problem in Machine Learning. *Highlights in Science, Engineering and Technolog ACME*, 111, 353–359.