SciencePG
Science Publishing Group

Research Article

Death Events from Heart Failure Prediction Using Machine Learning Approach

Hosea Isaac Gungbias [ⓘ]**, Mulapnen Haruna Kassem**[*] [ⓘ]

Computer Science Department, Plateau State University, Bokkos, Nigeria

Abstract

Heart failure is a significant global health concern, contributing to high mortality rates and imposing substantial burdens on healthcare systems. Early prediction of mortality in heart failure patients can facilitate timely interventions, enhance patient management, and improve overall survival outcomes. This study applies machine learning techniques to predict death events among heart failure patients using clinical data. Five classification algorithms—Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), and Gaussian Naïve Bayes—are implemented on a dataset containing 5,000 patient records with 13 clinical attributes obtained from Kaggle. The research methodology includes extensive data preprocessing, such as missing value imputation using mean/mode strategies, standardization, feature selection via ANOVA P-value testing, and data balancing with the Synthetic Minority Over-sampling Technique (SMOTE). Model optimization was performed through hyperparameter tuning and cross-validation to enhance predictive accuracy. The results from two experimental settings—one without optimization and one with hyperparameter tuning, feature selection, and Principal Component Analysis (PCA)—show that K-Nearest Neighbor achieved the highest accuracy (98.5%) and precision (98.9%) after optimization. In contrast, Random Forest performed exceptionally well without tuning, achieving an accuracy of 99.2% and an F1-score of 98.7%. The findings demonstrate the effectiveness of machine learning in heart failure prognosis, providing valuable insights for clinical decision-making and personalized patient care.

Keywords

Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor, Gaussian Naïve Bayes, Heart Failure, Death Event

1. Introduction

Heart failure is a critical area of research due to its widespread prevalence and its profound effects on patient well-being and healthcare infrastructures [10]. According to the World Health Organization, cardiovascular diseases (CVD) claim approximately 17.9 million lives globally each year [8]. Given the high mortality rates, accurately predicting heart-related illnesses and failures based on clinical data from diverse patient records has become essential for healthcare professionals.

Machine learning algorithms have emerged as a powerful tool for analyzing complex medical datasets, enabling the prediction of critical health outcomes, such as death events in heart failure patients. This research seeks to utilize various machine learning models to forecast mortality risk in heart failure patients using comprehensive clinical records.

The dataset employed in this study encompasses a range of

clinical attributes from heart failure patients, including age, gender, serum creatinine levels, ejection fraction, and several other health markers. The primary objective is to build and assess classification models capable of accurately predicting death events in these patients. By employing sophisticated data analysis techniques, the study aims to enhance prediction accuracy, ultimately improving patient management and potentially saving lives.

2. Related Works

The prediction of death events related to heart diseases has been a key focus in medical research, with several researchers achieving notable results. In 2021, S. Saravanan and K. Swaminathan explored heart failure prediction using a hybrid approach that combined k-means clustering with a support vector machine (SVM). Their method involved grouping the dataset into six clusters using k-means, followed by applying the SVM algorithm to these clusters, yielding an impressive accuracy of 93.33%.

Further survival analysis on heart failure patients was carried out by P. Makam and G. Janardhan [8], utilizing a range of machine learning models, including Logistic Regression, Decision Tree, Random Forest, XGBoost, and AdaBoost. Their study reported an accuracy of 88.5% for Logistic Regression and 85.24% for both Decision Tree and XGBoost.

Additionally, Ravulapalli et al. [9] conducted a comprehensive evaluation of machine learning classifiers in predicting heart failure. Their study tested various classification techniques, including Stochastic Gradient Descent (SGD), Logistic Regression (LR), and Decision Tree-based models like AdaBoost, Support Vector Machine (SVM), and Random Forest (RF). The research involved comparing the performance of these algorithms on an imbalanced dataset derived from heart failure clinical data.

Recent advancements have further explored innovative techniques for heart failure mortality prediction. Li et al. [6] developed a machine learning model using the XGBoost algorithm to predict in-hospital mortality among critically ill patients with acute heart failure. Utilizing data from the MIMIC-IV database, their model demonstrated superior predictive ability compared to traditional scoring systems, effectively assisting clinicians in early intervention strategies.

Zhang et al. [14] investigated Random Forest and Deep Learning approaches for mortality risk assessment in heart failure patients, showing that deep learning models significantly improved prediction accuracy compared to traditional machine learning methods.

Shi et al. [11] explored preprocessing methods for imbalanced clinical datasets, particularly focusing on data scaling, outlier processing, and resampling techniques. Their study demonstrated that an optimized preprocessing pipeline could enhance one-month mortality prediction accuracy in heart failure patients.

Additionally, a study by Makam and Janardhan [7] examined the integration of voice-driven biomarkers with machine learning to enhance mortality prediction among hospitalized heart failure patients. Their findings revealed that voice biomarkers, when combined with traditional clinical features, improved predictive accuracy over conventional models.
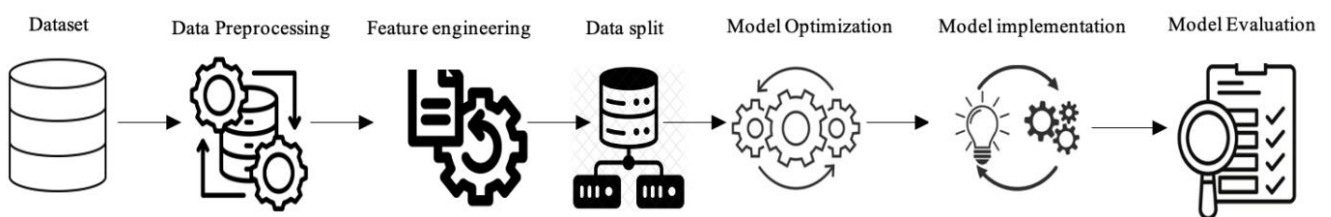
3. Methodology

Figure 1. Research Framework.

The figure above shows the framework of the research study from the dataset to the model evaluation.

3.1. Data Cleaning and Preprocessing

The dataset used for this analysis was sourced from Kaggle's Heart Failure Clinical Records, comprising 5,000 patient records with 13 features that include demographic, health, and behavioral data. The "DEATH_EVENT" column represents the occurrence of heart failure.

To enhance model performance, numerical features were standardized using the StandardScaler function, which adjusted the values to a mean of 0 and a standard deviation of 1. This standardization was applied after dividing the dataset into training and test sets to prevent data leakage.

Missing values were handled as follows:

Missing values in numerical columns were imputed with the median for the age column and the mean for the other numerical features. Figure 2 (below) illustrates the percentage of missing values in each column.

Missing values in categorical columns were filled in with

the mode.

These preprocessing steps helped to improve the dataset's

quality, reduce bias, and boost the performance of the machine learning models.
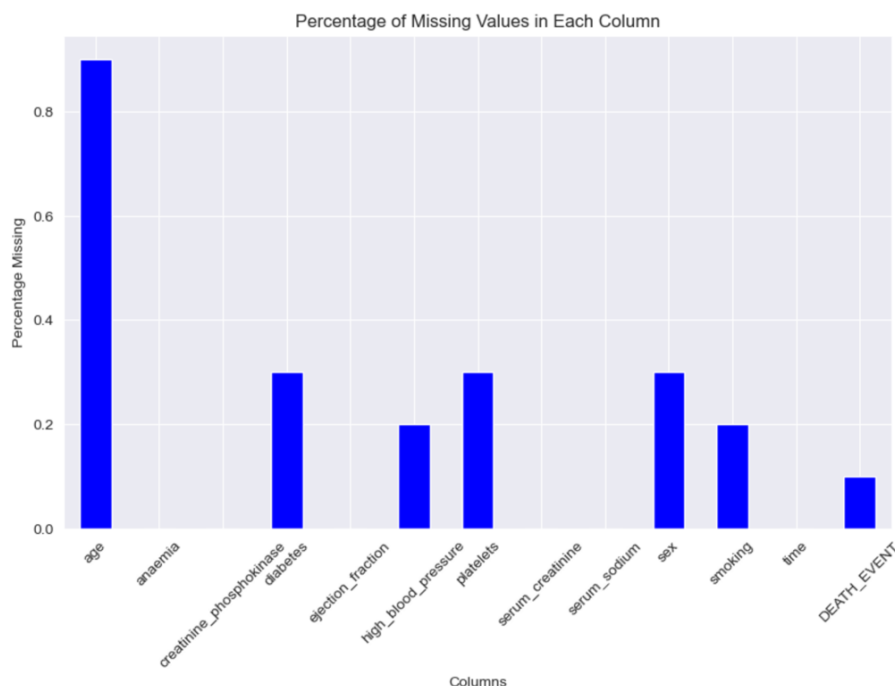

Figure 2. Percentage of missing values.

3.2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is instrumental in identifying patterns and trends within the data by highlighting its key characteristics. To conduct the analysis, essential libraries such as Seaborn, NumPy, and Pandas were imported, enabling tasks like value counts to assess the frequency of categorical variables. Figure 3 (below) demonstrates how the EDA reveals the dataset's imbalanced nature.
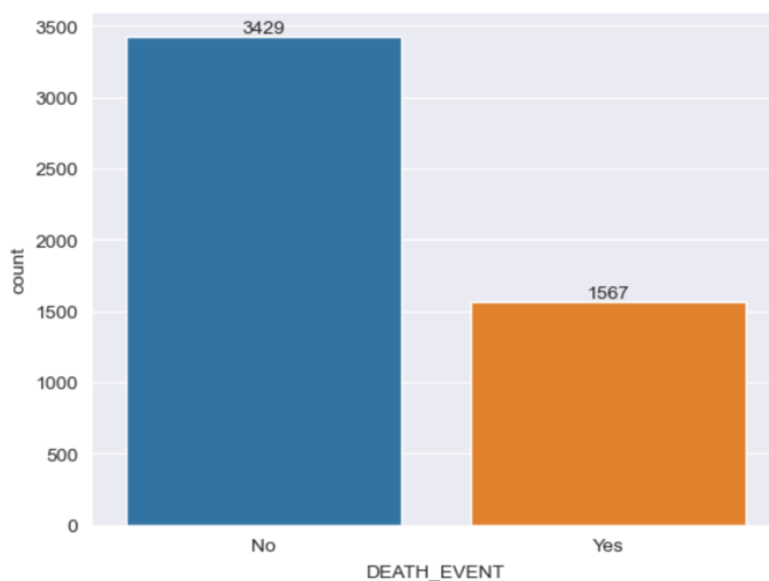

Figure 3. Target variable showing imbalanced data.

As shown in Figure 4, the correlation matrix displays how different features are related to each other. A negative value in the matrix represents an inverse relationship, meaning that as one variable increases, the other decreases. Conversely, a positive value signifies a direct relationship, where an increase in one variable corresponds with an increase in the related feature.
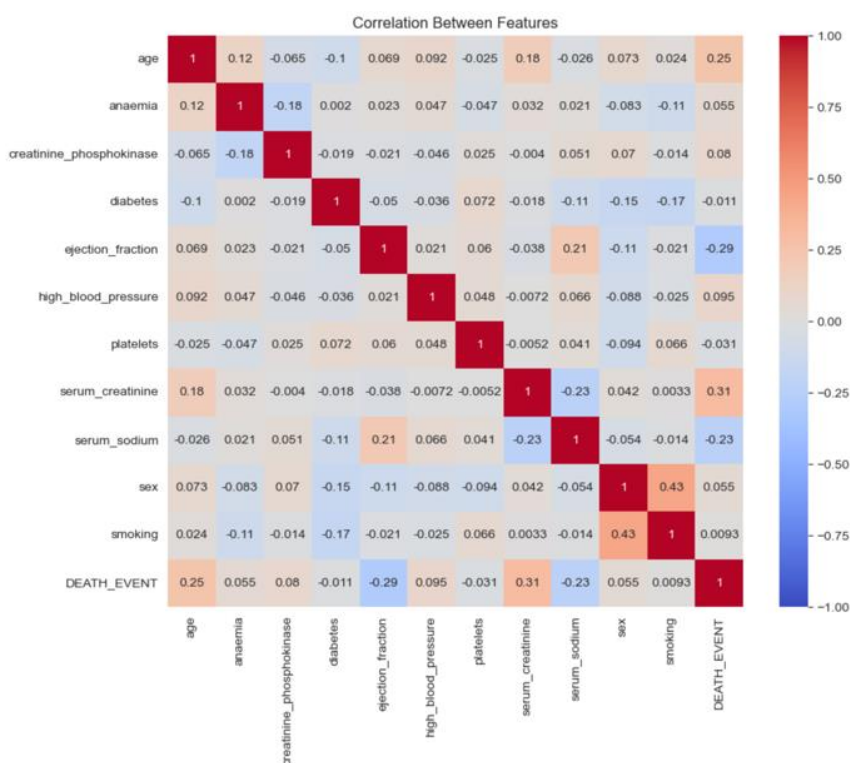

Figure 4. Correlation matrix between features.

Figure 5 depicts the distribution of categorical data within the dataset. Analyzing this distribution is essential for detecting imbalances or patterns that could impact the performance of machine learning models.
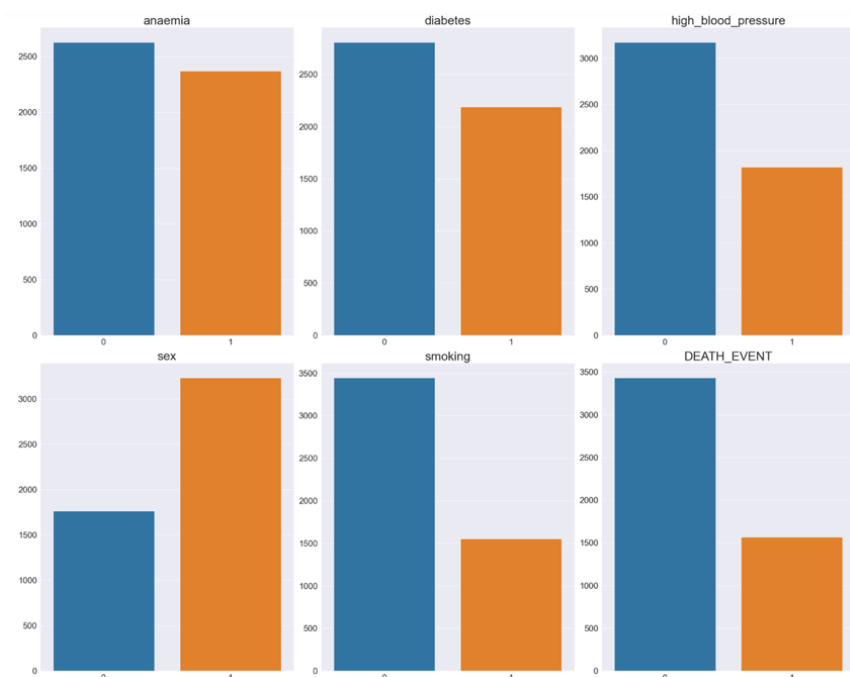

Figure 5. Distribution of Categorical data.

4

3.3. Feature Engineering

This research utilized various data preprocessing techniques to ready the dataset for machine learning models. These techniques included:

Encoding: Label encoding was applied to transform categorical variables into numerical values.

Data Split: The dataset was divided into 80% for training and 20% for testing, with a random state of 42 to ensure con-

sistent and reproducible results.

Feature Selection: This method is used to identify the features that are most statistically relevant to the prediction label. The ANOVA P-value test helps determine these significant features. A P-value below 0.05 suggests that the difference between group means is statistically significant, indicating that the feature has a strong influence on the prediction label (as illustrated in Figure 6).
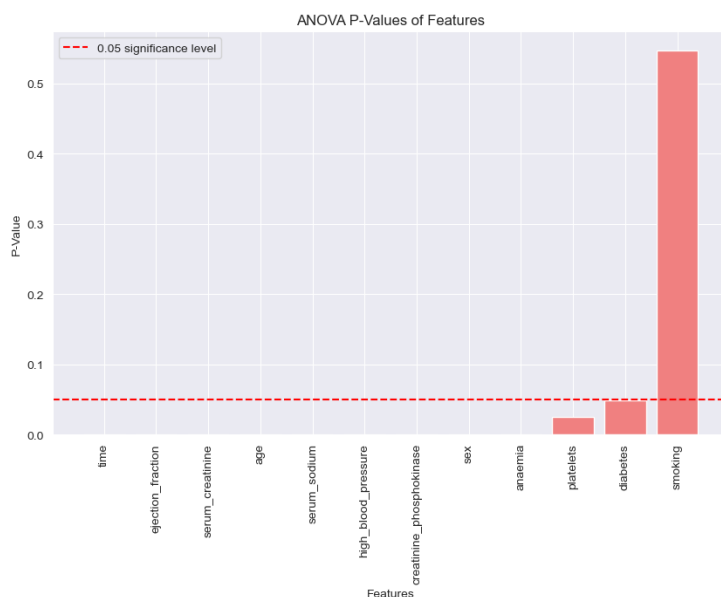
Figure 6. *Significant Features.*

Overfitting: To prevent the model from overfitting, where it learns the training data too closely and struggles to generalize to new data, it was crucial to implement measures. In this study, Principal Component Analysis (PCA) was used to reduce overfitting. Figure 7 below displays the PCA results for the dataset.
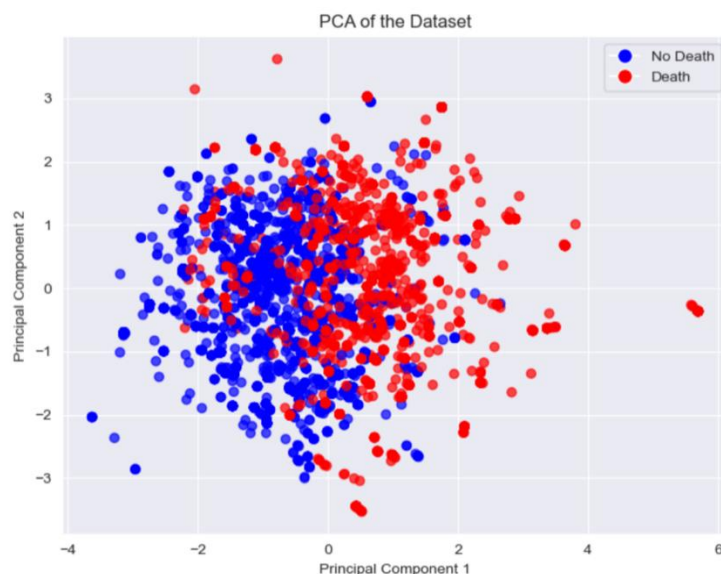
Figure 7. *PCA of the data.*

3.4. Model Optimization

Oversampling technique: The dataset is imbalanced, with a larger number of "NO" cases in the "DEATH_EVENT" category compared to "YES" cases (illustrated in Figure 8). This

imbalance poses a risk of bias in the model's performance. To mitigate this issue, the SMOTE oversampling technique was used, which entails choosing a random subset of the minority class (YES) and duplicating it to align with the size of the majority class (NO), thereby creating a balanced dataset.
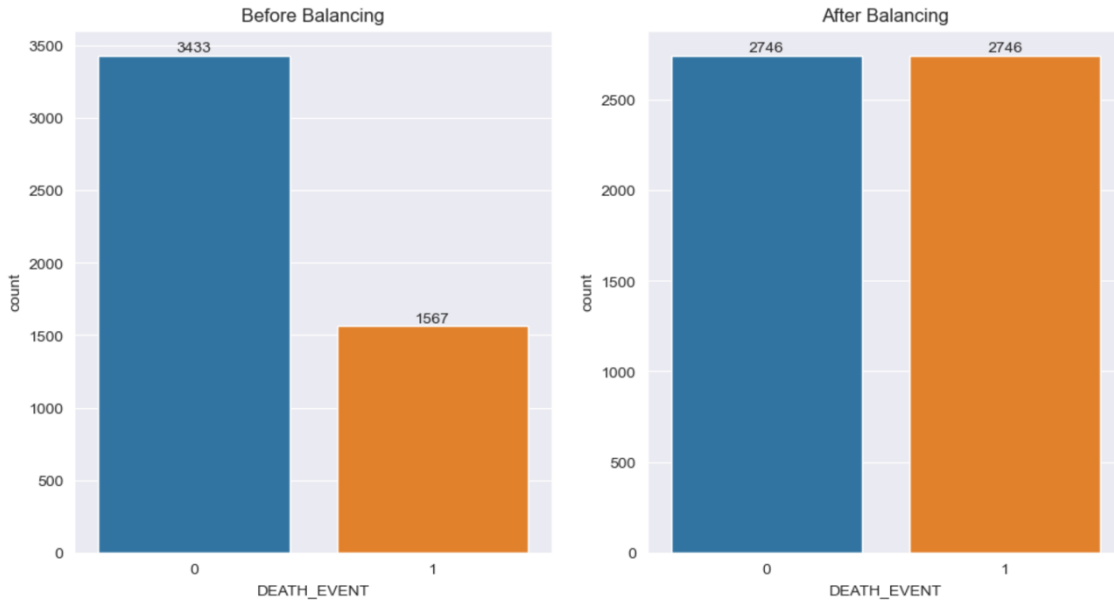

Figure 8. *Data balancing.*

To optimize model parameters, grid search was utilized to examine various hyperparameter values and identify the most effective combination, thereby improving performance. K-fold cross-validation with five folds was implemented to enhance the estimation of performance by training on different subsets and testing on the leftover data. This procedure was repeated five times for each fold to minimize variance and yield a more precise performance estimate. Following the optimization, the algorithm was executed again, and the enhanced performance was illustrated using a bar chart, confusion matrix, and ROC curve.

3.5. Model Implementation

This research utilized five different machine learning techniques: Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest Classifier (RF), Gaussian Naïve Bayes (GNB), and Logistic Regression (LR). All these algorithms are widely recognized and have been extensively applied in binary classification tasks over the years.

Logistic Regression: Logistic regression is a predictive model in machine learning that outputs values between 0 and 1. Also known as the logistic model, it estimates the likelihood of a specific event occurring. There are three types of logistic regression: binomial, multinomial, and ordinal. This model uses an S-shaped logistic function for its predictions

and is commonly applied in classification tasks [1, 5]. To achieve optimal performance, it is essential to minimize multicollinearity. Unlike linear regression, which utilizes a linear cost function, logistic regression applies a threshold value to categorize data into two distinct classes.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

The Logistic Regression function is shown in equation 1 above where; $e = base\ of\ nature\ logarithms\ and\ x = variable\ to\ be\ predicted$.

Decision Tree: This algorithm divides the data according to feature values, forming a binary classification framework resembling a tree [4]. Its straightforward nature and ease of visualization made it an ideal option for our project. It works particularly well with smaller datasets that have a limited number of features.

Random Forest Classifier: The Random Forest algorithm is a powerful machine learning method that constructs multiple decision trees during training. It uses the mode of the classes for classification tasks or the average prediction for regression tasks, drawing from the outputs of these individual trees. This approach employs an ensemble of decision trees formed through techniques such as bagging and random feature selection [14]. The algorithm improves the model's accuracy and mitigates overfitting by evaluating a subset of features at

each split, based on purity analysis. This leads to a robust model that performs effectively on both large and small datasets, providing enhanced accuracy and resilience to noise in the data.

Gaussian Naïve Bayes: The Gaussian Naïve Bayes classifier is based on the assumption that features are independent of one another, known as the independence assumption [12]. This simplification allows the model to be fast and efficient, particularly with high-dimensional datasets. However, this assumption can be limiting, as real-world data often shows correlations between features. When such correlations exist, the classifier's accuracy may decrease. Despite this drawback, it remains a favored option due to its simplicity, speed, and effectiveness in many practical scenarios, especially when the independence assumption holds true.

K-nearest Neighbor: The k-nearest neighbor (KNN) method is a traditional and widely utilized machine learning algorithm, especially favored for classification tasks. Its popularity stems from its simplicity and strong classification performance. Unlike many contemporary algorithms that rely on neural networks, KNN does not necessitate an extensive training period. Instead, it classifies a given instance by identifying the k-nearest data points and determining the majority class among those neighbors [13]. This straightforward approach has established KNN as a fundamental technique in the field of machine learning.

3.6. Model Evaluation

The model was evaluated using some metrics:

Accuracy: The proportion of cases that are accurately categorized [2]. Accuracy is given by; $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$.

F1 Score: F1-score combines both recall and precision therefore providing a single measure of quality [3]. F1-score is given by; $F1-score = 2 \times \frac{(Precision \times Recall)}{(Precison + Recall)}$.

Recall: Recall is the measure of correctly identified True Positives given by; $Recall = \frac{TP}{TP + FN}$.

Precision: Precision measures the True positive prediction out of all positive predictions. Precision is given by; $Precision = \frac{TP}{TP + FP}$.

4. Results and Discussions

Two experiments were conducted, and their results were compared.

Experiment One: The model was built without tuning the hyper-parameters.

Experiment Two: The hyper-parameters were tuned, ANOVA P-value feature selection was applied and PCA was applied before building the model.

Table 1. Result showing performance of experiment one and experiment two.

	Experiment One				Experiment Two			
	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
Logistic Regression	81	72.2	79.2	66.3	82	81.6	81.2	82
Naïve Bayes	84	73.5	68.4	79.6	83	82.2	79.7	84.8
Decision Tree	98.7	97.9	98.4	97.5	97.5	97.4	96.6	98.3
Random Forest	99.2	98.7	99	98.4	98.5	98.5	98.4	98.5
K-nearest Neighbor	97.5	96	97.1	95	98.5	98.5	98	98.9

The performance results presented in Table 1 compare the models from the first and second experiments. In the second experiment, algorithms like Naïve Bayes and KNN exhibited improved performance. Conversely, Logistic Regression demonstrated consistent performance across both experiments. In contrast, Decision Tree and Random Forest algorithms performed better in the first experiment.
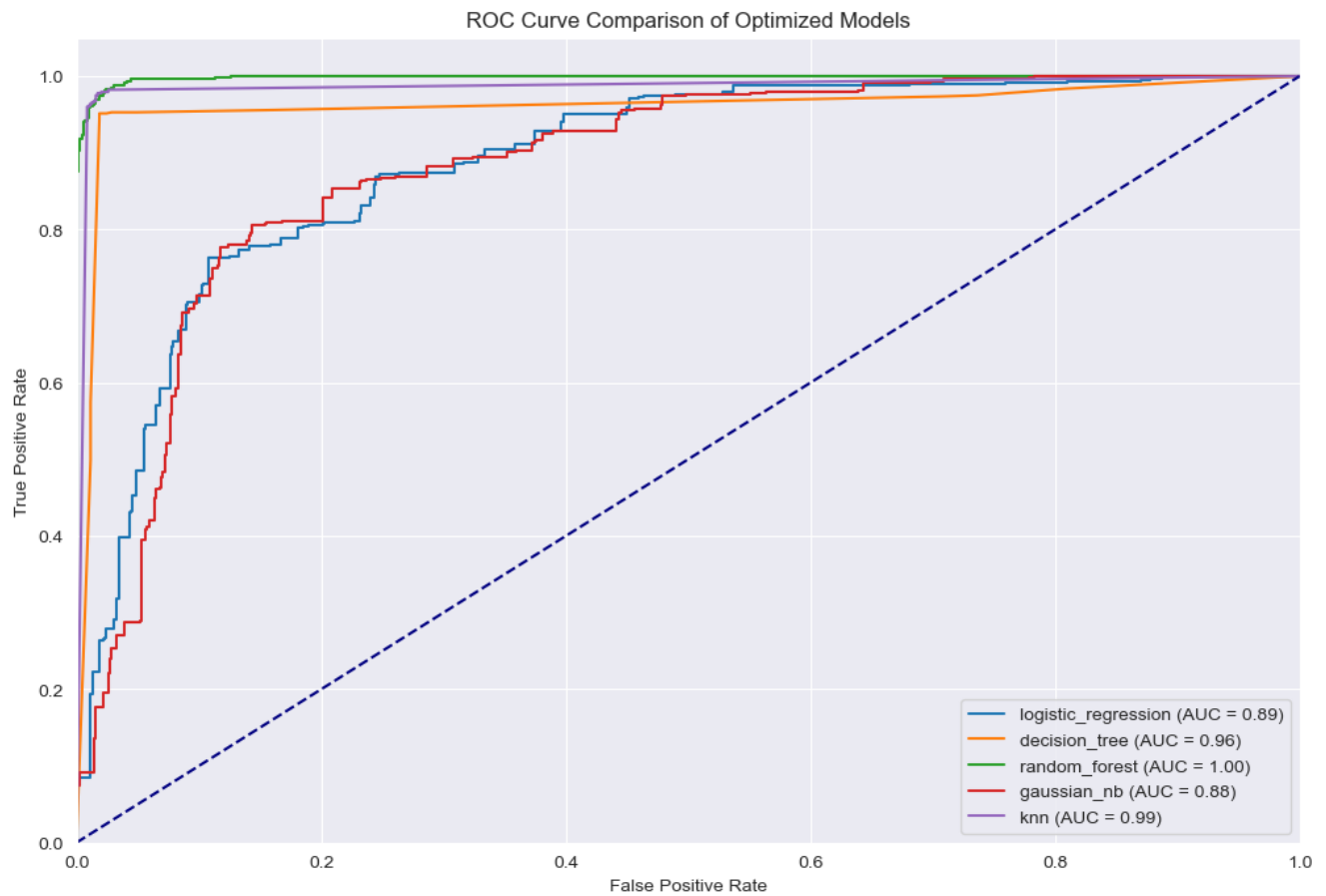
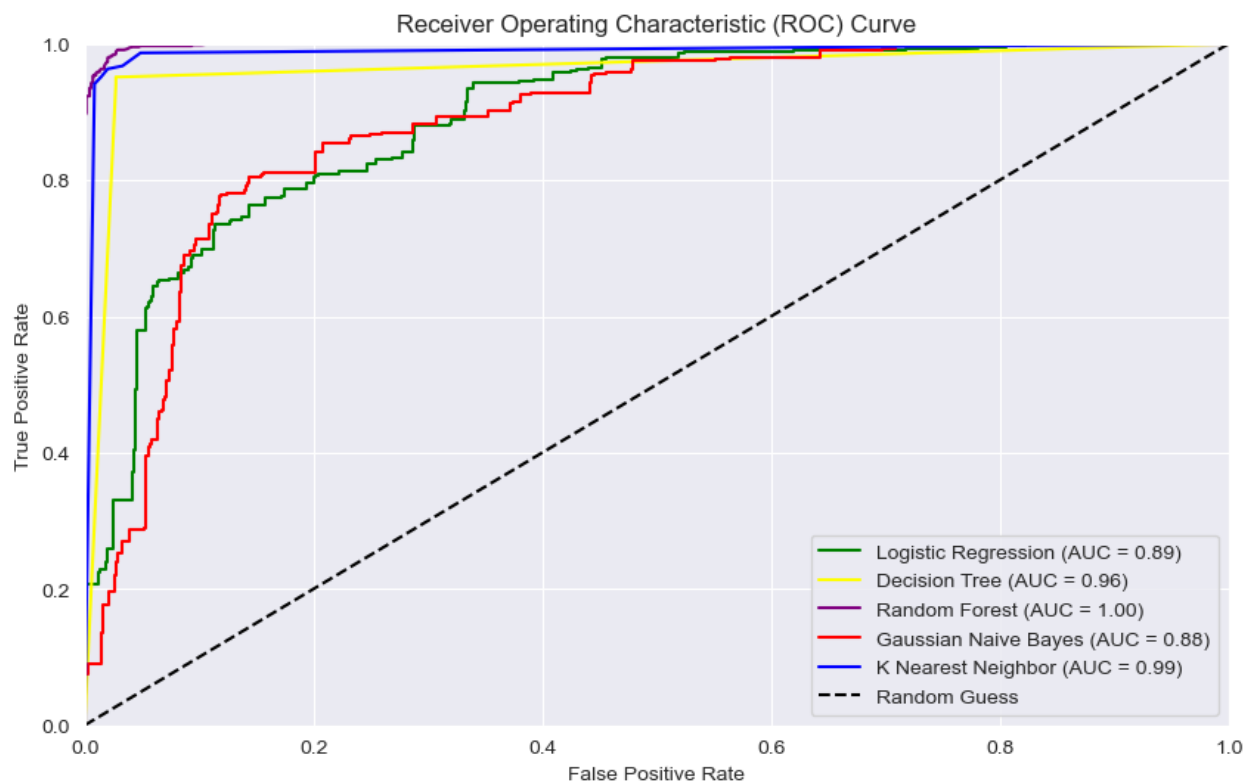*Figure 9. Experiment One ROC-Curve.*

Figure 10. Experiment Two ROC-Curve.

Figures 8 and 9 reveal that the ROC curve remained unchanged across both experiments. This suggests that the models' capability to differentiate between positive and negative classes was consistent, regardless of the application of hyper-parameter tuning, ANOVA feature selection, or PCA. The similarity of the ROC curves indicates that these additional steps did not notably affect the overall discriminatory power of the models in this instance.

5. Conclusion

In this study, a range of techniques was investigated to improve the performance of predictive models for heart failure mortality events. The research involved two experiments: the first was conducted without any optimization, while the second incorporated hyper-parameter tuning, ANOVA feature selection, and Principal Component Analysis (PCA). The objective was to evaluate how these optimization strategies influenced the performance of the models. The results revealed that all algorithms benefitted from the applied optimizations; however, the Random Forest algorithm exhibited only a slight increase in precision compared to its performance without optimizations. Furthermore, the ROC curves from both experiments showed no significant change in the models' ability to discriminate between positive and negative outcomes. These findings highlight the necessity of implementing tailored optimization strategies for different algorithms to attain the best possible model performance. It suggests that while optimizations can enhance certain aspects of model accuracy, their impact may vary across different methods, indicating the need for a nuanced approach to model optimization in predictive analytics.

6. Future Works

Future research work can aim to;
1) Investigate other feature selection methods like Recursive Feature Elimination (RFE).
2) Employ advanced hyper-parameter tuning techniques such as Bayesian optimization.
3) Explore ensemble learning methods like stacking or blending to improve accuracy.
4) Validate the models using external datasets to ensure robustness and generalizability.

Abbreviations

KNN	K-Nearest Neighbor
PCA	Principal Component Analysis
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic

Author Contributions

Hosea Isaac Gungbias: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing

Mulapnen Haruna Kassem: Investigation

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Bekhet, H. A. and Eletter, S. F. K. (2014) 'Credit risk assessment model for Jordanian commercial banks: Neural scoring approach', Review of Development Finance, 4(1), pp. 20-28. Available at: https://doi.org/10.1016/j.rdf.2014.03.002

[2] Hosea, I. G. et al. (2023) 'A Machine Learning Approach to Fake News Detection Using Support Vector Machine (SVM) and Unsupervised Learning Model', Advances in Multidisciplinary and Scientific Research Journal Publication, 11(1), pp. 11. Available at:
https://doi.org/10.22624/aims/csean-smart2023p1

[3] J. Sepúlveda and S. A. Velastın (2015) 'F1 score assesment of gaussian mixture background subtraction algorithms using the MuHAVi dataset', 6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15). Available at: https://doi.org/10.1049/ic.2015.0106

[4] Joshi, G. (2022) 'Distributed Optimization in Machine Learning', in 'Distributed Optimization in Machine Learning', pp. 1-12.

[5] K. Yadav and S. Singh (2023) 'Loan Status Prediction using SVM and Logistic Regression', 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). Available at:
https://doi.org/10.1109/ICCCNT56998.2023.10307473

[6] Li, J., Sun, Y., Ren, J., Wu, Y., & He, Z. (2024). "Machine Learning for In-hospital Mortality Prediction in Critically Ill Patients With Acute Heart Failure: A Retrospective Analysis Based on the MIMIC-IV Database." *Journal of Cardiothoracic and Vascular Anesthesia*. Available at:
https://pubmed.ncbi.nlm.nih.gov/39779429/

[7] Makam, P., & Janardhan, G. (2024). "Voice-Driven Mortality Prediction in Hospitalized Heart Failure Patients: A Machine Learning Approach Enhanced with Diagnostic Biomarkers." *arXiv preprint arXiv: 2402. 13812*. Available at:
https://arxiv.org/abs/2402.13812

[8] P. Makam and G. Janardhan (2023) 'Survival Analysis of Heart Failure Patients Using Advanced Machine Learning Techniques', 2023 International Conference on Advanced & Global Engineering Challenges (AGEC). Available at:
https://doi.org/10.1109/AGEC57922.2023.00029

[9]   Ravulapalli, L. T. et al. (2023) 'Evaluative Study of Machine Learning Classifiers in Predicting Heart Failure: A Focus on Imbalanced Datasets', Ingénierie Des Systèmes D'Information, 28(3), pp. 717-724. Available at: https://doi.org/10.18280/isi.280322

[10]   S. Jamal, W. A. Elenin and L. Chen (2023) 'Developing and Evaluating Data-Driven Heart Disease Prediction Models by Ensemble Methods on Different Data Mining Tools', 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). Available at: https://doi.org/10.1109/UEMCON59035.2023.10315997

[11]   Shi, Y., et al. (2023). "Enhancing Mortality Prediction in Heart Failure Patients: Exploring Preprocessing Methods for Imbalanced Clinical Datasets." *arXiv preprint arXiv: 2310. 00457*. Available at: https://arxiv.org/abs/2310.00457

[12]   Y. Shi et al. (2021) 'Efficient Jamming Identification in Wireless Communication: Using Small Sample Data Driven Naive Bayes Classifier', IEEE Wireless Communications Letters, 10(7), pp. 1375-1379. Available at: https://doi.org/10.1109/LWC.2021.3064843

[13]   Z. Fan et al. (2024) 'Multiview Adaptive K-Nearest Neighbor Classification', IEEE Transactions on Artificial Intelligence, 5(3), pp. 1221-1234. Available at: https://doi.org/10.1109/TAI.2023.3296092

[14]   Zhang, T., et al. (2024). "Random Forest and Deep Learning Approaches for Mortality Risk Assessment in Heart Failure Patients." *Computational and Mathematical Methods in Medicine*, 2024, Article ID 1354827. Available at: https://doi.org/10.1155/2024/1354827