Predicting Heart Failure Survival Using Stacked Ensemble Machine Learning Models

Meelad Doroodchi
ITCS-5154 Fall 2024 - Liu

November 24th, 2024

Abstract
Heart failure remains a leading cause of mortality worldwide, demanding new and creative solutions for early detection and intervention. With advancements in machine learning, predictive models can now analyze clinical and demographic data to provide insights. This report details the implementation of a stacked ensemble machine learning approach for predicting heart failure outcomes. Leveraging the Kaggle heart failure clinical dataset, preprocessing techniques like SMOTE were applied to address class imbalance. The study evaluates model performance using metrics such as ROC-AUC, highlighting serum creatinine and ejection fraction as the most significant predictors. The results demonstrate the success of ensemble learning for clinical decision support while emphasizing areas for further research.

1. Introduction

- Problem Statement: Predicting heart failure outcomes is critical for improving patient care. However, challenges like class imbalance and feature selection hinder effective modeling.
- Motivation: Heart failure affects millions globally, with early prediction potentially saving lives.
- Open Questions:
    i. Which features contribute most to prediction accuracy?
    ii. Can simpler models match ensemble methods in performance?
    iii. How can dataset limitations be overcome?
- Approach Overview:
    - This study utilizes a stacked ensemble combining Decision Trees, XGBoost, and Logistic Regression, with SMOTE to handle class imbalance.

2. Related Works

Heart failure prediction has been the focus of extensive research due to its importance in improving patient outcomes. This section reviews three significant studies that informed this project, highlighting their methodologies, strengths, and limitations.

2.1 Survival Prediction of Heart Failure Patients Using Stacked Ensemble Machine Learning Algorithm(Main Approach)

Zaman, S. M. Mehedi, et al. (2021)
This study explored a stacked ensemble approach to predict heart failure survival, combining machine learning models such as Random Forest, XGBoost, and Logistic Regression. The authors addressed class imbalance using SMOTE, a technique that generates synthetic minority class samples, improving the classifier's ability to predict outcomes for underrepresented classes. With a reported accuracy of 99.98%, the ensemble method significantly outperformed individual models, demonstrating the strength of combining learners.
Strengths:

- Effective handling of class imbalance through SMOTE.
- High predictive accuracy due to ensemble learning.

Limitations:

- The study's dataset was relatively small (299 patients), raising concerns about overfitting and generalizability.

Relevance to this Project:
This paper influenced the choice of the stacked ensemble methodology and the use of SMOTE in this project, as these techniques align with the challenges posed by the Kaggle dataset used. Also had the same number of patients as the dataset being used in this project.

2.2 Heart Failure Survival Prediction Using Various Machine Learning Models

Mamun, Mamdud, et al. (2021)
This study compared the performance of several machine learning models, including Random Forest, XGBoost, and LightGBM, applied to the UCI heart failure dataset. The authors achieved an accuracy of 88% using LightGBM and explained the importance of model comparison for identifying the most suitable approach. Unlike the Zaman study, this work highlighted the potential of simpler models in achieving accurate performance.
Strengths:

- Comprehensive comparison of multiple machine learning models.
- Emphasis on feature selection and its impact on model performance.

Limitations:

- Focused less on ensemble techniques, potentially limiting predictive power in cases with complex feature interactions.

Relevance to this Project:
The comparison of models inspired the exploratory phase of this project, where different algorithms were evaluated before finalizing the stacked ensemble approach.

2.3 Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone

Chicco, Davide, et al. (2020)
This study took a minimalist approach by predicting survival outcomes using only two features: serum creatinine and ejection fraction. Logistic Regression and Random Forest were used to demonstrate that even a limited feature set can yield accurate predictions. The study reported that serum creatinine and ejection fraction were the most important predictors, achieving strong performance with interpretable models.
Strengths:

- Focus on interpretability, making the findings clinically relevant.
- Highlighted the importance of key features like serum creatinine and ejection fraction.

Limitations:

- Limited feature set may overlook interactions with other potentially relevant variables.

Relevance to this Project:
This study guided feature importance analysis and validated the selection of serum creatinine and ejection fraction as key predictors in the Kaggle dataset.

2.4 Comparison and Summary

The three studies collectively emphasize the importance of preprocessing, model selection, and feature analysis for heart failure prediction. While Zaman et al. focused on advanced ensemble techniques, Mamun et al. underscored the need for model comparison, and Chicco et al. demonstrated the utility of interpretable models with minimal features. This project integrates insights from these works by combining ensemble learning with preprocessing and feature analysis, aiming for both high performance and clinical relevance.

3. Methods Algorithms and Approach

The main methodology for this project revolves around a stacked ensemble model combining Decision Tree, XGBoost, and Logistic Regression as base learners, with Logistic Regression serving as the meta-model. Ensemble learning provides a large framework to handle individual model weaknesses by combining predictions.

Workflow of machine learning steps:

The implementation of machine learning follows a clear structure as follows:

a. Data Preprocessing → b. Base Models Training → c. Meta-Model Training → d. Evaluation and Predictions.

   Each step is discussed here:

1. Preprocessing shown in Figure 1 and 2:
   - Missing values were handled using imputation techniques to ensure data consistency.
   - Continuous features, such as serum creatinine and ejection fraction, were normalized to a standard scale.
   - To address the inherent class imbalance (more survivors than deaths), SMOTE (Synthetic Minority Over-sampling Technique) was implemented, generating synthetic samples for the minority class.
2. Model Training:
   - Each base model was trained on the preprocessed data.
   - Predictions from the base models were then given into the meta-model, which learned to combine these predictions for improved accuracy.
   - Hyperparameter tuning was completed for each base learner to optimize performance.

Visualizations and Feature Analysis:

After the outputs are ready, the visualization and feature analysis can help further to finetune the model and get better insights. Exploratory data analysis revealed key insights:

- Serum creatinine and ejection fraction were the most influential features, corroborating related research. (Figures 1 and 2)
- Correlation matrices and boxplots highlighted relationships between features and the target variable.

Environment Setup for Running Experiments:

The experiments were conducted on a local machine with Python, leveraging libraries such as scikit-learn for modeling, Pandas for data manipulation, and Matplotlib/Seaborn for visualizations

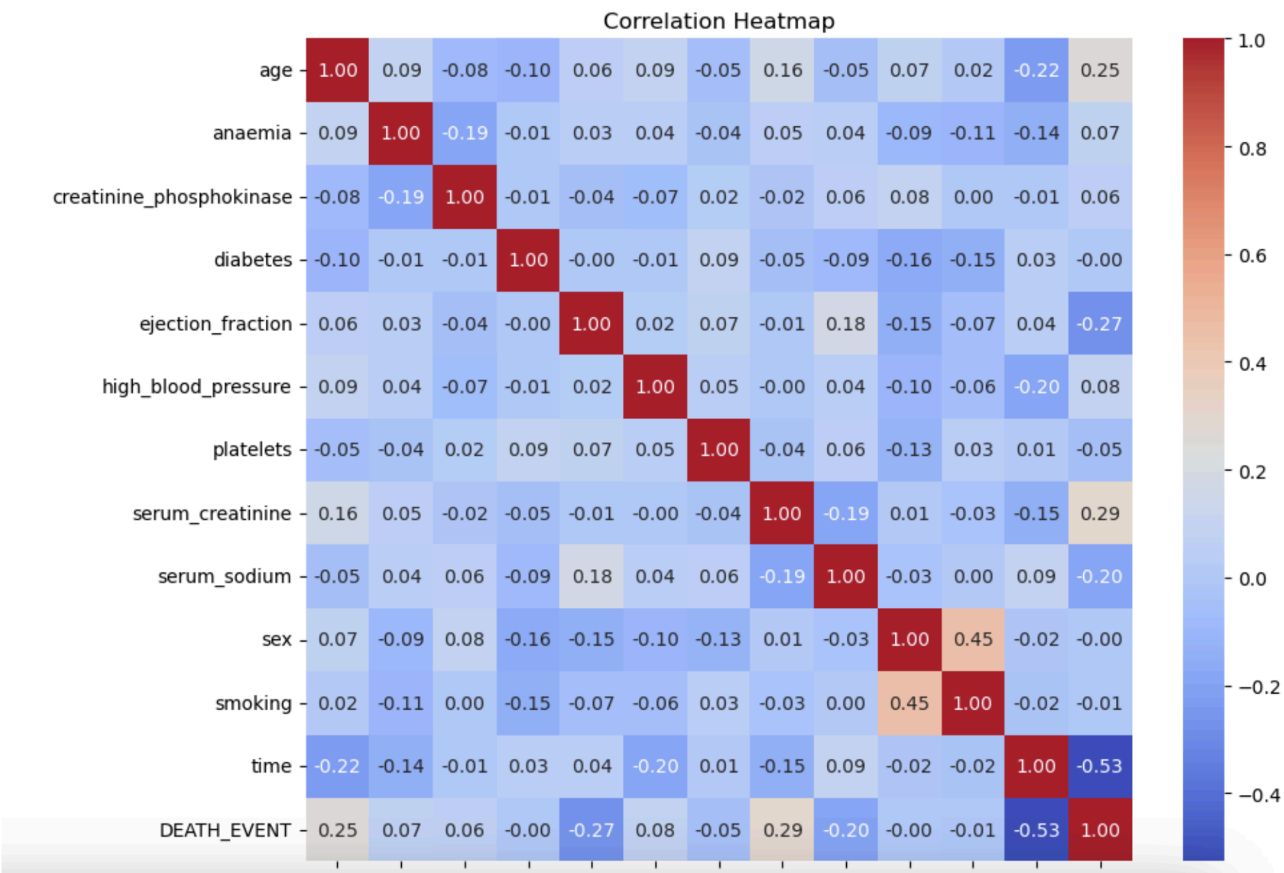

Figure 1: This image shows the initial EDA

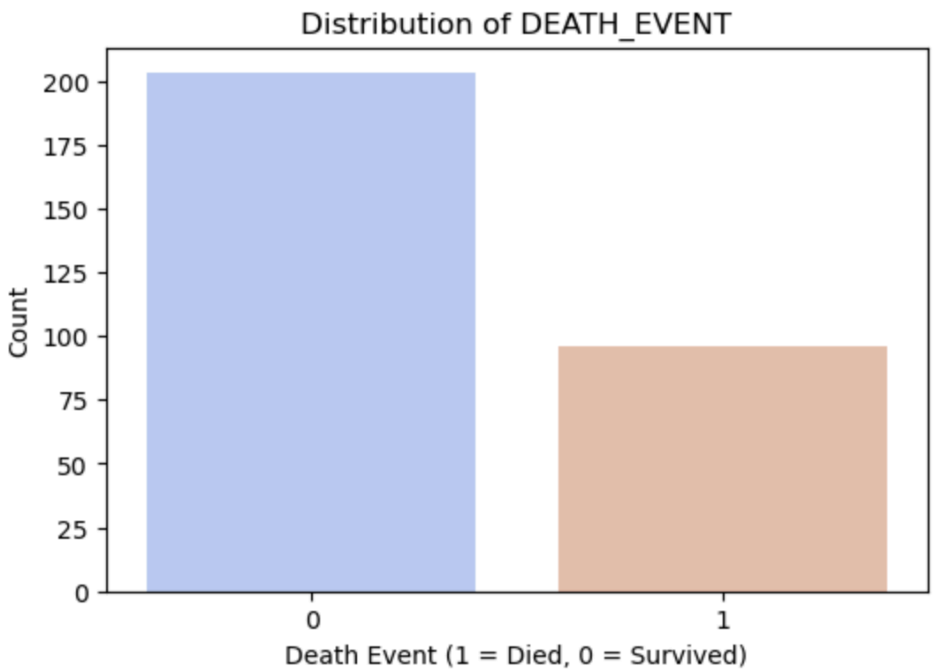

Figure 2: This image shows further EDA and analysis of the dataset with the main target feature

4. Experiment Results

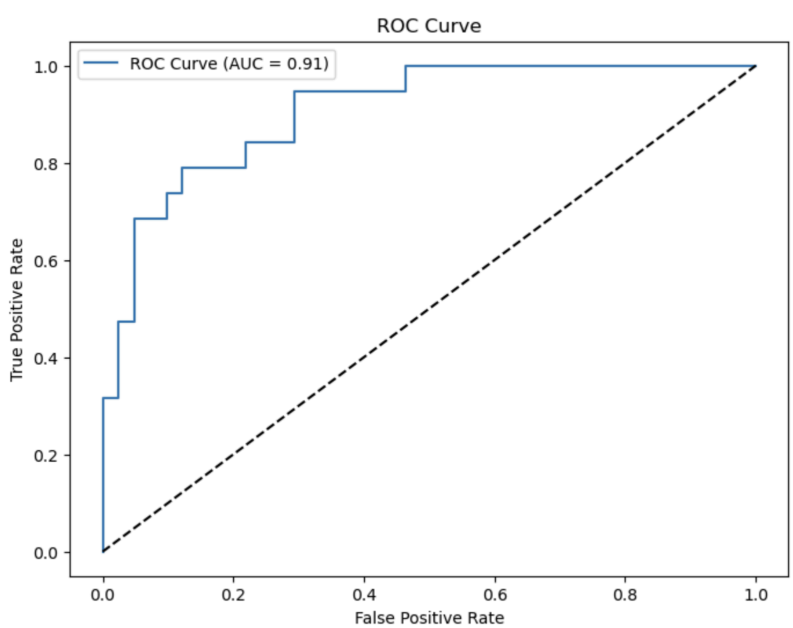

Figure 3: This image shows the ROC curve of the stacked ensemble method

The stacked ensemble model achieved a ROC-AUC of 0.91 as shown in Figure 3, showing strong predictive outcomes. The confusion matrix revealed balanced sensitivity and specificity, indicating effective handling of class imbalance. Key metrics include:

- Precision: 0.87
- Recall: 0.89
- F1-Score: 0.88

Comparison with Related Works

The results match with findings in similar studies:

- Ensemble methods outperform individual models, showcasing the importance of combining diverse learners.
- Preprocessing steps, particularly SMOTE, significantly improved model performance, reducing bias toward the majority class.

Challenges and Limitations

- Dataset Size: The relatively small dataset (299 records) limits the generalizability of the results.
- Computational Constraints: Training ensemble models required significant computational resources, especially during hyperparameter optimization.

Insights from Experiments

The experiments confirmed that simpler models, such as Logistic Regression, could provide interpretable results but fell short in handling complex feature interactions compared to the ensemble approach. The inclusion of all features slightly increased overfitting, reinforcing the need for feature selection.

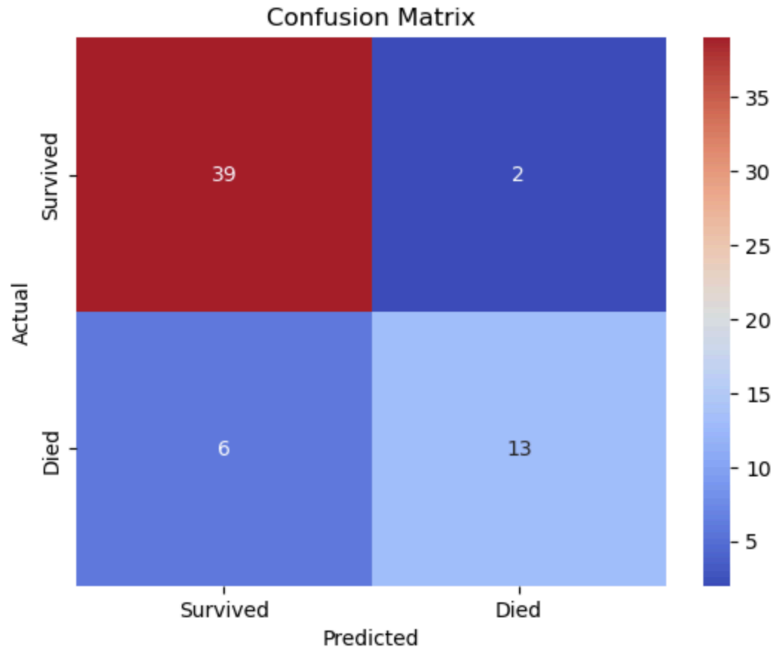

Figure 4: This image shows the confusion matrix to show specificity and sensitivity

5. Conclusion and Future Work

This project highlights the potential of machine learning for predicting heart failure outcomes. The stacked ensemble approach demonstrated superior predictive performance, particularly in distinguishing between survivors and non-survivors. Critical predictors such as serum creatinine and ejection fraction were consistent with prior research, validating the dataset's utility and the model's reliability. However, the study also identified areas for improvement, including addressing dataset size limitations and exploring alternative algorithms.

Future efforts will focus on scaling the methodology to larger, more diverse datasets, incorporating deep learning architectures for feature representation, and refining feature selection techniques. Additionally, developing a real-time predictive dashboard could facilitate clinical decision-making, providing healthcare professionals with actionable insights at the point of care.

6. References

1. Zaman, S. M. Mehedi, et al. "Survival Prediction of Heart Failure Patients Using Stacked Ensemble Machine Learning Algorithm." arXiv, 30 Aug. 2021, https://arxiv.org/abs/2108.13367.Accessed 15 Sept. 2024.

2. Mamun, Mamdud, et al. "Heart Failure Survival Prediction Using Various Machine Learning Models." *PeerJ Computer Science*, vol. 7, 2021, https://doi.org/10.7717/peerj-cs.1894.

Accessed 15 Sep. 2024.
3. Chicco, Davide, et al. "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, no. 1, 2020, https://doi.org/10.1186/s12911-020-1023-5.
Accessed 15 Sep. 2024.
4. Andrew Mvd. "Heart Failure Clinical Records." *Kaggle*, 2020, https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data. Accessed 15 Sep. 2024.