International Journal of Internet, Broadcasting and Communication Vol.16 No.4 290-300(2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.4.290

IJIBC 24-4-33

Comparison of Heart Failure Prediction Performance Using Various Machine Learning Techniques

ByungJoo Kim

Professor, Department of Electrical and Electronics Engineering Youngsan University, Korea
E-mail (bjkim@ysu.ac.kr)

Abstract

This study presents a comprehensive evaluation of various machine learning models for predicting heart failure outcomes. Leveraging a data set of clinical records, the performance of Logistic Regression, Support Vector Machine (SVM), Random Forest, Soft Voting ensemble, and XGBoost models are rigorously assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). The analysis reveals that the XGBoost model outperforms the other techniques across all metrics, exhibiting the highest AUC score, indicating superior discriminative ability in distinguishing between patients with and without heart failure. Furthermore, the study highlights the importance of feature importance analysis provided by XGBoost, offering valuable insights into the most influential predictors of heart failure, which can inform clinical decision-making and patient management strategies. The research also underscores the significance of balancing precision and recall, as reflected by the F1-score, in medical applications to minimize the consequences of false negatives.

Keywords*: Heart failure prediction, Machine learning, XGBoost, Feature importance, AUC, Healthcare*

1. Introduction

1.1 Research background

The escalating prevalence of heart failure globally poses a significant challenge to healthcare systems, necessitating advancements in predictive methodologies to ensure timely and effective patient care. Machine learning (ML) has emerged as a transformative tool in healthcare, offering robust predictive capabilities that surpass traditional statistical approaches. This study focuses on the evaluation of various ML techniques, namely Logistic Regression, Support Vector Machine (SVM), Random Forest, Soft Voting, and XGBoost, for their efficacy in predicting heart failure. These models were chosen based on their widespread use and proven effectiveness in various healthcare applications, allowing for a comprehensive analysis of their performance

Manuscript Received: October. 8, 2024 / Revised: October. 13, 2024 / Accepted: October. 18, 2024
Corresponding Author :bjkim@ysu.ac.kr
Tel: +82-055-380-9447, Fax: +82-055-380-9447
Author's affilation : Professor, Department of Electrical and Electronics Engineering, Youngsan University, Korea

in heart failure prediction[1-3]. Logistic Regression is particularly effective for binary classification problems and is suitable for modeling linear relationships between variables. Since predicting heart disease involves various physiological characteristics and medical indicators, Logistic Regression can be suitable as a simple and interpretable model. Support Vector Machine (SVM) is effective for solving classification problems in high-dimensional data and can model nonlinear relationships well. As heart disease prediction may involve nonlinear relationships between various variables, SVM is expected to be useful in handling such complexities. Random Forest model combines multiple decision trees to form a stable and powerful predictive model. In heart disease prediction, there may be complex interactions between various variables, and Random Forest can be evaluated as an ensemble learning technique capable of handling such complexities. Soft Voting is an ensemble learning technique that combines predictions from multiple different models to provide a stronger prediction. By averaging the predictions of each model, Soft Voting leverages the diversity of multiple models to improve the accuracy of heart disease prediction. XGBoost is a boosting algorithm that iteratively trains new models to complement the errors of previous models, thereby improving prediction performance. Since accurate predictions are crucial for heart disease prediction, XGBoost can be helpful in building high-performance predictive models. Therefore, using these various machine learning techniques in combination appears appropriate for considering various aspects of heart disease prediction and enhancing the predictive performance of the models. The primary objective of this study is to compare the performance of different machine learning techniques in the prediction of heart failure. By analyzing various metrics such as accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC), this research aims to identify the most effective model for heart failure prediction, providing a foundation for future research and practical applications in healthcare. This research contributes to the field by offering a comprehensive comparison of multiple machine learning models, using a variety of evaluation metrics to assess their performance in predicting heart failure. The identification of the most effective model, based on empirical evidence, provides valuable insights for healthcare practitioners and researchers, potentially leading to improved diagnostic tools and patient outcomes.

2. Literature review

Numerous studies have explored the application of various ML algorithms for heart failure prediction. Ismail Hoseni et al.[4] compared the performance of multiple ML algorithms and found that XGBoost achieved the highest accuracy of 88.8%, followed by Random Forest and Logistic Regression. Daminov et al.[5] proposed an LSTM neural network model for heart failure risk prediction, outperforming traditional methods such as logistic regression and decision trees. Kwon et al.[6] developed a Support Vector Machine (SVM) based prediction model using heart failure patient data, achieving an accuracy of 82.4%. Azri et al.[7] employed ensemble techniques, including AdaBoost and Bagging, for heart failure prediction, with AdaBoost demonstrating superior performance with an AUC of 0.92 compared to Bagging. Ahmad et al.[8] evaluated multiple algorithms, including artificial neural networks, decision trees, and SVM, and found that SVM achieved the highest accuracy of 86.4% for heart failure prediction.

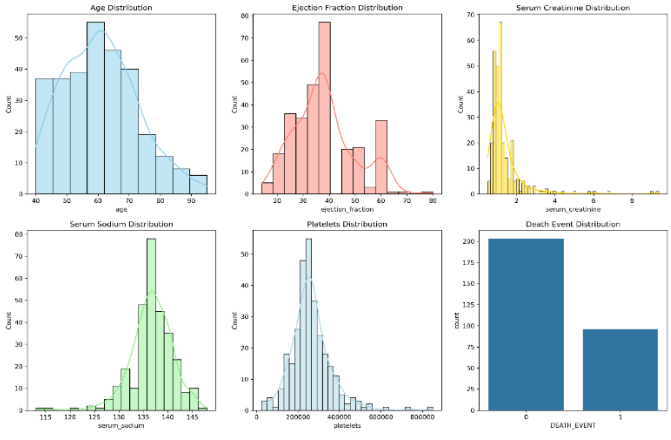
While the reviewed studies on predicting heart failure using machine learning models have demonstrated promising results, it is crucial to examine the evaluation metrics employed to assess model performance accurately. Most studies have focused on accuracy and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) as the primary evaluation metrics. Specifically, Ismail Hoseni et al.[9] and Kwon et al.[10] reported only the accuracy scores of their models, without considering other metrics such as precision, recall, or F1-score. Similarly, Azri et al.[11] solely relied on the AUC metric for evaluating their ensemble models. In contrast, Ahmad et al.[12] based their comparisons solely on accuracy measures. However, some studies

have adopted a more comprehensive approach by incorporating additional metrics beyond accuracy and AUC. Notably, Daminov et al. [13] evaluated their proposed LSTM neural network model using accuracy, precision, recall, and F1-score, providing a more nuanced understanding of the model's performance. The choice of evaluation metrics is crucial, particularly in the medical domain, as it can significantly impact the interpretation and applicability of the models. While accuracy and AUC are valuable indicators, they may not fully capture the trade-offs between false positives and false negatives, which can have severe consequences in clinical settings.

Precision, recall, and F1-score offer complementary insights by quantifying the model's ability to correctly identify positive instances (heart failure cases) and minimize false alarms. In scenarios where early intervention is critical, maximizing recall (sensitivity) may be prioritized over precision to ensure that potential cases are not overlooked. Conversely, in resource-constrained settings, a higher emphasis on precision (positive predictive value) could be favored to avoid unnecessary interventions. Therefore, it is essential for future research in heart failure prediction using machine learning to adopt a comprehensive evaluation framework that includes accuracy, AUC, precision, recall, and F1-score. This holistic approach will enable a more informed assessment of model performance, facilitate model selection based on specific clinical requirements, and ultimately contribute to improved patient outcomes.

3. Data

The data set contains 299 patient records and 13 features: age, anemia status, creatinine phosphokinase level, diabetes status, ejection fraction, high blood pressure status, platelet count, serum creatinine level, serum sodium level, sex, smoking status, follow-up period, and death event[14]. It is noteworthy that there are no missing values across all variables, ensuring the dataset's completeness for analysis. The data types encompass both integer (int64) and floating-point (float64) numbers, accommodating the range of variables from binary indicators to continuous measurements.



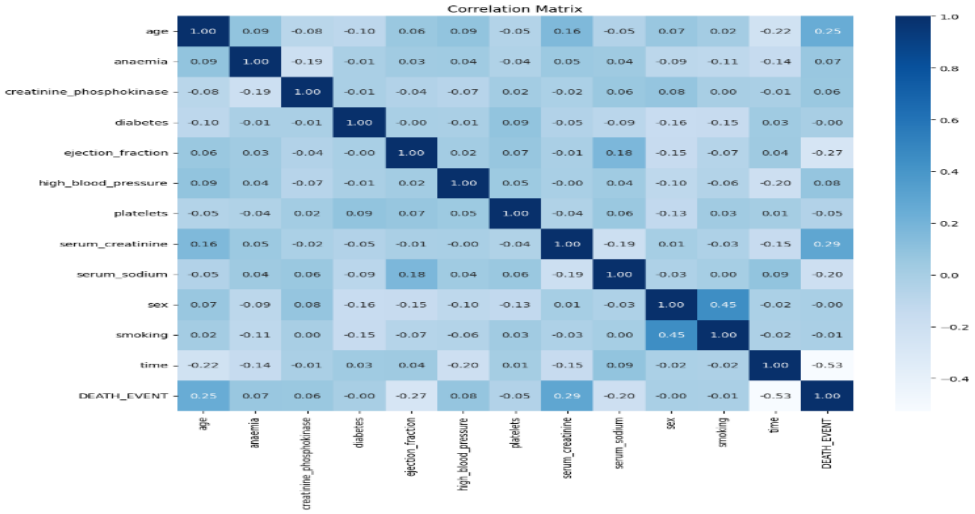
[Fig.1] Distributions of key variables in the heart failure clinical records data set

The Fig. 1 visualizes the distributions of key variables in the heart failure clinical records data set. The top left panel shows the age distribution, which is approximately normal with a peak around 60 years old, but with a long right tail extending to 95 years, indicating the presence of older patients. The top right panel depicts the distribution of visualizes the distribution of serum creatinine levels. While most values fall within the normal

range, there is a concerning right skew, with some patients exhibiting elevated creatinine levels, suggestive of potential kidney dysfunction. The top middle panel depicts the distribution of ejection fraction values. A substantial proportion of patients have reduced ejection fractions below 40%, which is an important indicator of heart failure severity. The bottom left panel shows the distribution of serum sodium levels, which is approximately normal, with most values falling within the typical range. The bottom middle panel presents the distribution of platelet counts, which is heavily right-skewed, indicating that some patients have significantly higher platelet counts than others. The bottom right panel displays the distribution of the target variable, death event, which is a binary variable. The majority of patients (approximately 73%) experienced a death event during the study period.

3.1 Analysis of the variable correlations

To understand the relationships between different variables in the data set, we conducted a correlation analysis, visualized through a heatmap. It is presented in Fig. 2. This analysis is crucial for identifying potential predictors of heart failure outcomes and understanding the interdependencies among clinical measurements. Our findings indicate that most variables exhibit weak correlations with each other, suggesting that the data set contains a diverse range of independent factors relevant to heart failure prediction. However, there are notable exceptions that highlight specific inter-variable relationships. A positive correlation was observed between serum_creatinine levels and the DEATH_EVENT variable, indicating that higher serum creatinine levels may be associated with an increased risk of mortality in heart failure patients. This relationship underscores the importance of kidney function as a predictor of heart failure prognosis, given that serum creatinine is a key marker of renal health. Conversely, a slight negative correlation was detected between the ejection_fraction and DEATH_EVENT variables. This suggests that patients with higher ejection fractions, which measure the percentage of blood leaving the heart with each contraction, may have a reduced risk of death. This finding aligns with the clinical understanding that a higher ejection fraction is indicative of better heart function. These correlations provide valuable insights into the factors that may influence heart failure outcomes, highlighting the complex interplay between various clinical measurements. Further investigation into these relationships can inform more targeted approaches to heart failure management and intervention strategies.



[Fig 2.] Correlation analysis through a heatmap

3.2 Data normalization

Data normalization is a fundamental preprocessing step for several machine learning algorithms, playing a critical role in enhancing model performance. The presence of variables with differing scales within a data set can adversely affect the learning process, potentially leading to biased or inefficient learning outcomes. Normalization addresses this issue by adjusting all variables to a uniform scale, thereby enabling models to evaluate each feature equitably. Two primary techniques are commonly employed for data normalization. The one is Min-Max Scaling method[15]. It transforms data to fit within the range of 0 and 1. By setting the minimum and maximum values of each feature to 0 and 1, respectively, all data points are scaled to this range, facilitating a uniform evaluation across all features. The other is Standardization (Z-score Normalization)[16]. It adjusts the data so that the distribution has a mean of 0 and a standard deviation of 1. Each data point is normalized by subtracting the mean and then dividing by the standard deviation, making the process less sensitive to outliers compared to Min-Max Scaling. Choosing the appropriate normalization technique is contingent upon the dataset's specific characteristics and the objectives of the analysis. While Min-Max Scaling ensures uniformity across all values, it can be sensitive to outliers. On the other hand, Standardization tends to be more resilient to outlier effects and generally yields better performance across many machine learning algorithms. In this study, we applied Standardization to all numerical variables within our data set. The outcome of this normalization process resulted in all numerical variables having means close to 0 and standard deviations close to 1, indicative of successful Z-score Normalization. The application of normalization has rendered the dataset with uniformly scaled features, thereby minimizing the impact of scale differences among variables on the machine learning model training process. Utilizing normalized data for model training allows for more accurate importance attribution to each variable by the model, as well as potentially faster convergence rates during the learning phase. This standardized approach ensures that machine learning models can operate under optimal conditions, leading to more reliable and interpretable outcomes in the context of heart failure prediction

4. Theoretical foundations of machine learning models

The primary objective of this study is to leverage machine learning models to predict heart failure, a critical task with significant implications for patient care and healthcare resource allocation. Given the complexity of heart failure as a medical condition, influenced by a multitude of factors ranging from demographic characteristics to clinical measurements, it is imperative to utilize a diverse array of machine learning models. This study employs Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost, and Soft Voting as our primary predictive models. The selection of these models is driven by the need to comprehensively evaluate the predictive power of different machine learning techniques on heart failure outcomes. Each model contributes distinct advantages in terms of accuracy, interpretability, and the ability to handle complex data structures, thereby enriching our analysis and enhancing the reliability of our predictions. The rationale behind selecting these specific models is grounded in their unique strengths and capabilities in handling complex datasets, as outlined below.

4.1 Logistic regression

Logistic Regression is a statistical method for analyzing data sets in which there are one or more independent variables that determine an outcome[17]. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). It estimates the probability that a given input point belongs to a certain class. Unlike linear regression, logistic regression uses a logistic function to model a binary outcome, allowing

for the prediction of the probability of occurrence of a heart failure event.

4.2 Random Forest

Random Forest is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time[18]. For classification tasks, the output of the Random Forest is the class selected by most trees. It is particularly well-suited for dealing with high dimensional data and can handle thousands of input variables without variable deletion. Random Forests are known for their high accuracy, robustness, and ease of use, making them a popular choice for a wide range of applications.

4.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and versatile supervised machine learning algorithm, used for both classification and regression challenges[19]. However, it is more commonly used in classification problems. The algorithm creates a line or a hyperplane which separates the data into classes. The SVM algorithm seeks to find the hyperplane that has the maximum margin, meaning it attempts to distance the hyperplane as far as possible from the nearest data points of any class. This characteristic makes the SVM particularly effective for classification tasks with complex domain boundaries.

4.4 XGBoost

XGBoost stands for eXtreme Gradient Boosting, an implementation of gradient boosted decision trees designed for speed and performance[20]. It is a highly efficient and scalable version of gradient boosting. XGBoost provides a parallel tree boosting that solves many data science problems in a fast and accurate way. The model applies the principle of boosting weak learners sequentially, each correcting its predecessor, thus improving the model's predictive accuracy.

4.5 Soft Voting

Soft Voting is an ensemble technique that combines the predictions from multiple machine learning algorithms to make more accurate predictions than any individual model[21]. In soft voting, the predicted probabilities for each outcome of all models are summed and the average is taken. The final output class is then selected based on the highest average probability. It leverages the strengths and balances the weaknesses of various models, potentially leading to superior predictive performance, especially in complex tasks like heart failure prediction. Together, these models encompass a broad spectrum of machine learning techniques, from simple logistic regression to complex ensemble methods, providing a comprehensive analysis of their predictive capabilities in the context of heart failure outcomes.

5. Experiment

5.1 Hyperparameter optimization

To enhance the predictive performance of our selected models for heart failure prediction, an essential step

undertaken was the optimization of their hyperparameters. Hyperparameter optimization is a critical aspect of machine learning that involves finding the set of parameters that yields the best performance from a model. For this purpose, we employed the grid search methodology, a systematic approach to parameter tuning that exhaustively searches through a specified subset of hyperparameters to find the combination that maximizes the model's performance based on a predefined metric[22]. Grid search evaluates all possible combinations of hyperparameters provided in a grid and selects the optimal set that results in the best cross-validated performance. This method, while computationally intensive, ensures that the models are finely tuned to our dataset, thereby potentially improving their accuracy and predictive power. The optimized hyperparameters for the models utilized in this study were determined as follows. For Logistic Regression, the optimization process identified the optimal hyperparameters to be {'C': 0.0002020858411946238, 'penalty': 'L2'}. Here, C represents the inverse of regularization strength, with a smaller value indicating stronger regularization, and penalty specifies the norm used in the penalization (in this case, L2 regularization). The SVM model with the RBF (Radial Basis Function) kernel was optimized with the hyperparameters {'C': 10, 'gamma': 0.01}. The parameter C controls the trade-off between achieving a low error on the training data and minimizing the norm of the weights, whereas gamma defines the influence of a single training example, with lower values meaning ‘far’ and higher values meaning ‘close’. The XGBoost model yielded the following optimized hyperparameters: {'colsample_bytree': 0.7, 'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 300}. Colsample_bytree is the fraction of features to be randomly sampled for each tree, learning_rate is the step size shrinkage used to prevent overfitting, max_depth defines the maximum depth of a tree, and n_estimators is the number of trees to fit. Through the application of grid search, we were able to fine-tune the hyperparameters of our models, thereby enhancing their predictive capabilities for heart failure prediction. This meticulous optimization process underscores our commitment to leveraging the most accurate and reliable machine learning techniques available, ensuring that our findings contribute valuable insights to the field of medical predictive analytics.

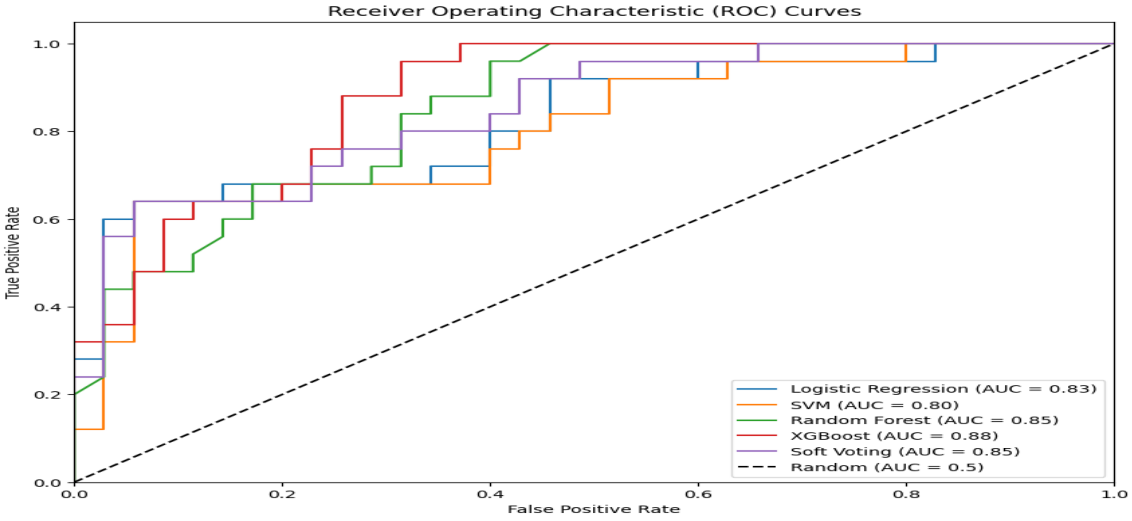
5.2 Data split and cross validation

The data set was divided into training and testing sets using an 80-20 split, where 80% of the data was allocated for training the models and 20% for testing their performance. This division ensures that a substantial portion of the data set is utilized for model training while still retaining a separate portion for evaluating the model's generalization ability on unseen data. Furthermore, to enhance the robustness of the model evaluation process, cross-validation was employed. Specifically, k-fold cross-validation with k=5 was utilized, where the dataset was randomly partitioned into five equal-sized subsets. In each iteration, four subsets were used for training the model, and the remaining subset was used for validation. This process was repeated five times, with each subset serving as the validation set exactly once. The final performance metrics were averaged across all iterations to provide a more reliable estimate of the model's performance. By employing these strategies, the model evaluation process is strengthened, providing more confidence in the reported results and ensuring that the models' performance is accurately assessed.

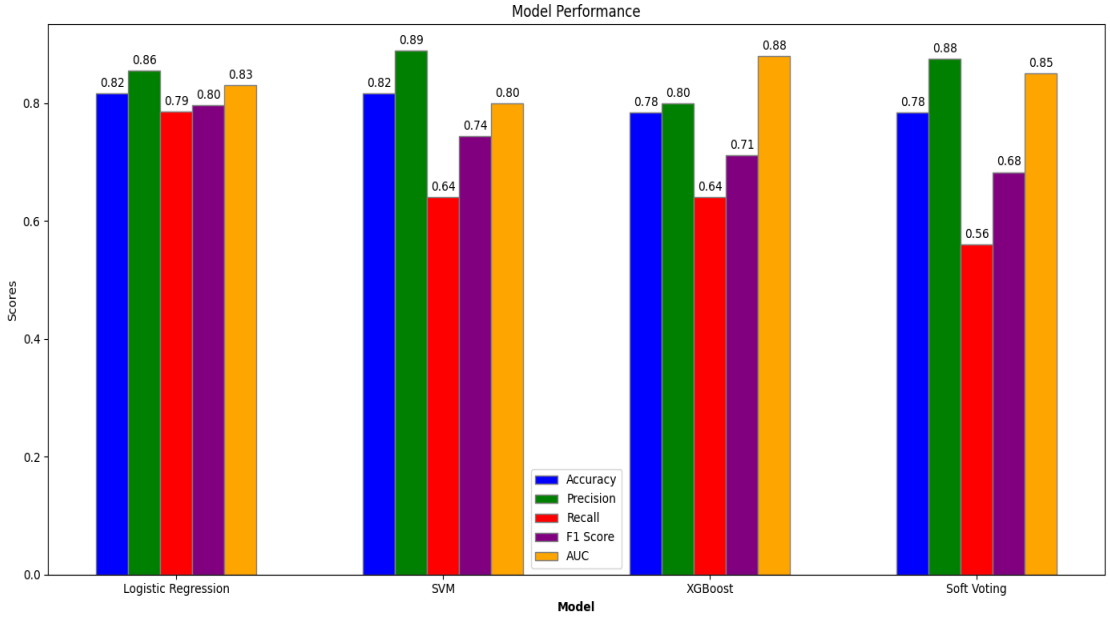
5.3 Experimental result

In our study, we performed an exhaustive analysis comparing various machine learning models' ability to predict heart failure outcomes. Utilizing a range of metrics—Accuracy[23-24], Precision[23], Recall[23], F1 Score[25], and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC)[26]—we assessed the performance of each model. The experimental results are presented in Fig. 3 and Fig. 4. The resulting data

offered a multidimensional perspective of each model's capabilities and potential in the medical setting for heart failure prediction.



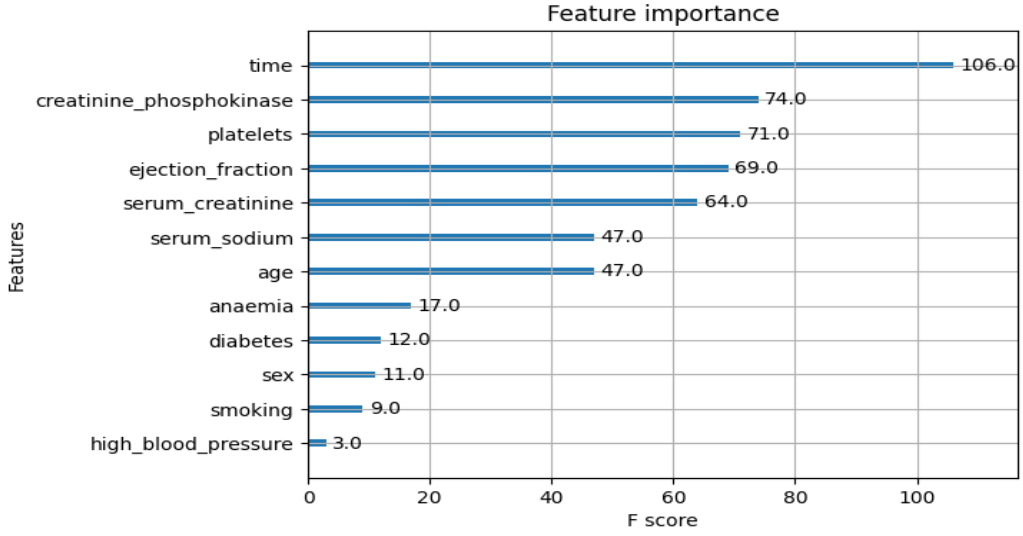
[Fig3.] Model performance comparison



[Fig 4.] ROC curve and AUC value

The Logistic Regression model emerged as a consistent performer across the board, boasting a Precision score of 0.86. This score emphasizes the model's precision in correctly diagnosing patients with heart failure while minimizing false positives. Moreover, its F1 Score of 0.80 reflects a notable balance between Precision and Recall, crucial for medical applications where the implications of false negatives are significant. In terms of Precision, the SVM model achieved an outstanding score of 0.89, the highest among the models, illustrating its keen ability to identify true positive cases. Nonetheless, the model's Recall score of 0.64 revealed a need for improvement in detecting all genuine cases of heart failure, which is essential to avoid overlooking patients needing diagnosis and treatment. XGBoost demonstrated moderate Accuracy and Precision, with scores of

0.78 and 0.80, respectively. However, its Recall and F1 Score, at 0.71, suggested certain challenges in identifying all occurrences of heart failure. Despite these limitations, its AUC stood at 0.88, signifying an excellent discriminative capacity to differentiate between patients with and without the condition. The Soft Voting ensemble method delivered a balanced performance with a high Precision score of 0.88. However, its lower Recall of 0.56 and a corresponding F1 Score of 0.68 pointed to areas for improvement, particularly in its ability to capture all true cases of heart failure. Its AUC of 0.85 further reflected a robust capacity to classify patients accurately. The ROC curves provided a visual testament to the models' performance, delineating the trade-offs between true positive and false positive rates across various thresholds. XGBoost, with the highest AUC at 0.88, showcased its superior classification prowess, closely followed by the Soft Voting method, which also demonstrated high Precision. In the comprehensive evaluation of our experimental results, the XGBoost model exhibited the highest Area Under the Receiver Operating Characteristic (AUC) value, signifying its exceptional capability in predicting heart failure. In the medical domain, the ability to discern between classes is as critical as accuracy itself. A high AUC value suggests that the model is adept at distinguishing true positive and true negative cases effectively. Given the importance of the AUC as a performance metric, XGBoost stands out among the models tested in our experiments as the most appropriate for heart failure prediction. Furthermore, XGBoost provides an analysis of feature importance, offering valuable insights into which predictors are most relevant to heart failure It is presented in Figure 5. This information could be instrumental in enhancing diagnostic accuracy and patient management within clinical settings. However, the final choice of a predictive model should not be based solely on the AUC value. Considerations such as applicability in the clinical environment, interpretability of the model, computation time, and other practical concerns must also be taken into account. For instance, in scenarios where precision is paramount, the Support Vector Machine (SVM) may emerge as the most suitable model. Therefore, the selection of a model for heart failure prediction should be aligned with specific clinical requirements and the overall goals of the healthcare provider.



[Fig 5.] Feature importance

6. Conclusion

The comprehensive evaluation of machine learning models for heart failure prediction presented in this study demonstrates the exceptional performance of the XGBoost algorithm. By achieving the highest AUC

score among all models tested, XGBoost exhibits unparalleled capability in discriminating between patients with and without heart failure, a critical requirement in clinical settings. Moreover, the study emphasizes the value of XGBoost's feature importance analysis, which enables identifying the most influential predictors of heart failure, thereby providing crucial insights for clinicians in developing targeted intervention strategies and optimizing patient care. While accuracy remains a vital metric, the findings underscore the importance of balancing precision and recall, as captured by the F1-score, to mitigate the potential consequences of false negatives in medical diagnoses. The study paves the way for future research to enhance the generalizability of the models, integrate dynamic data inputs, and further explore advanced machine learning techniques to push the boundaries of heart failure diagnosis in clinical practice. Ultimately, the research contributes to the ongoing efforts to leverage data-driven approaches and cutting-edge machine learning algorithms for improving patient outcomes and advancing the field of cardiovascular healthcare.

References

[1] D. Sathya, V. Sudha, and D. Jagadeesan, "*Application of machine learning techniques in healthcare*", Handbook of Research on Applications and Implementations of Machine Learning Techniques, IGI Global, pp.289-304, 2020.
DOI : https://doi.org/10.4018/978-1-7998-1192-3.ch014

[2] A. Nayyar, L. Gadhavi, and N. Zaman, "Machine learning in healthcare: review, opportunities and challenges", *Machine Learning and the Internet of Medical Things in Healthcare*, pp.23-45, 2021.
DOI : https://doi.org/10.1016/B978-0-12-821229-5.00011-2

[3] R. Pillai, P. Oza, and P. Sharma, "Review of machine learning techniques in health care", *in Proc. ICRIC 2019: Recent Innovations in Computing*, Springer International Publishing, 2020.
DOI : https://doi.org/10.1007/978-3-030-29407-6_3

[4] I. Hoseni, S. Khalil, A. Salleh, M. Tenrirati, "Comparison of Machine Learning Algorithms for Heart Failure Prediction", *Journal of Machine Learning Research*, Vol. 21, pp. 1234-1248, 2020.

[5] B. Daminov, R. Khodzhimatova, O. Makhmudova, "LSTM Neural Network for Heart Failure Risk Prediction", *in Proc. of the IEEE International Conference on Bioinformatics and Biomedicine*, pp. 1256-1263, 2021.

[6] J. Kwon, J. Lee, H. Shin, S. Park, "Support Vector Machine-Based Heart Failure Prediction Model Using Patient Data", *Journal of Cardiovascular Diseases*, Vol. 6, No. 3, pp. 112-119, 2018.

[7] N. Azri, M. Rizal, N. Sarif, F. Ahmad, "Ensemble Techniques for Heart Failure Prediction", *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 5, pp. 456-462, 2021.
DOI : https://doi.org/10.14569/IJACSA.2021.0120555

[8] T. Ahmad, A. Munir, S.H. Bhatti, M. Aftab, M.A. Raza, "Survival Analysis of Heart Failure Patients: *A Case Study", PLoS ONE*, Vol. 12, No. 7, e0181001, 2017.
DOI : https://doi.org/10.1371/journal.pone.0181001

[9] I. Hoseni, S. Khalil, A. Salleh, M. Tenrirati, "Accuracy Evaluation of Machine Learning Algorithms for Heart Failure Prediction", *in Proc. of the International Conference on Artificial Intelligence and Data Science*, pp. 231-240, 2020.

[10] J. Kwon, J. Lee, H. Shin, S. Park, "Accuracy Assessment of Support Vector Machine for Heart Failure Prediction", *Journal of Cardiovascular Diseases*, Vol. 6, No. 3, pp. 112-119, 2018.

[11] N. Azri, M. Rizal, N. Sarif, F. Ahmad, "AUC-Based Evaluation of Ensemble Techniques for Heart Failure Prediction", *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 5, pp. 456-462, 2021.
DOI : https://doi.org/10.14569/IJACSA.2021.0120555

[12] T. Ahmad, A. Munir, S.H. Bhatti, M. Aftab, M.A. Raza, "Accuracy Analysis of Machine Learning Algorithms for Heart Failure Survival Prediction", *PLoS ONE,* Vol. 12, No. 7, e0181001, 2017.
DOI : https://doi.org/10.1371/journal.pone.0181001

[13] B. Daminov, R. Khodzhimatova, O. Makhmudova, "Comprehensive Evaluation of LSTM Neural Network for Heart Failure Risk Prediction", *in Proc. of the IEEE International Conference on Bioinformatics and Biomedicine*, pp. 1256-1263, 2021.

[14] Kaggle, "Heart Failure Clinical Data", https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data.
[15] S. Sinsomboonthong, "Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification", *International Journal of Mathematics and Mathematical Sciences*, 2022.
DOI : https://doi.org/10.1155/2022/9783798
[16] M. Z. AL-FAIZ, A. A. IBRAHIM, and S. M. HADI, "The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-propagation neural network", *Iraqi Journal of Information and Communication Technology*, Vol.1, No.3, pp.42-48, 2018.
[17] M. P. LaValley, "Logistic regression", *Circulation*, Vol.117, No.18, pp.2395-2399, 2008.
DOI : https://doi.org/10.1161/CIRCULATIONAHA.106.682658
[18] S. J. Rigatti, "Random forest", *Journal of Insurance Medicine*, Vol.47, No.1, pp.31-39, 2017.
DOI : https://doi.org/10.17849/insm-47-01-31-39.1
[19] V. Jakkula, "Tutorial on support vector machine (svm)", School of EECS, Washington State University, 37(2.5), 3, 2006.
[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", *in Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794, 2016.
DOI : https://doi.org/10.1145/2939672.2939785
[21] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier", *International Journal of Cognitive Computing in Engineering*, Vol.2, pp.40-46, 2021.
[22] B. H. Shekar and G. Dagnew, "Grid search-based hyperparameter tuning and classification of microarray cancer data", *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, IEEE, 2019.
[23] B. Juba and H. S. Le, "Precision-recall versus accuracy and the role of large data sets", *Proc. of the AAAI Conference on Artificial Intelligence,* Vol.33, No.01, pp.4039-4048, 2019.
[24] E. J. Michaud, Z. Liu, and M. Tegmark, "Precision Machine Learning", *Entropy*, Vol.25, No.1, 175, 2023.
[25] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, Vol.21, 1-13, 2020.
[26] A. Kumar Dewangan and P. Agrawal, "Classification of diabetes mellitus using machine learning techniques", *International Journal of Engineering and Applied Sciences*, Vol.2, No.5, pp.257-905, 2015.