Volume 12 Issue 2 April 2026

IJRDO Journal of Applied Science
ISSN:2455-6653
Research Type

MACHINE LEARNING-BASED PREDICTION OF MORTALITY RISK IN HEART FAILURE PATIENTS USING CLINICAL DATA

[1*]Dr. John J.V. McMurray, [2]Dr. Mandeep R. Mehra

[1*]Professor of Medical Cardiology Institute of Cardiovascular and Medical Sciences University of Glasgow, Scotland
[2]Professor of Medicine / Cardiovascular Specialist Brigham and Women's Hospital

Dr. John J.V. McMurray, Email: murray.reseacher@gmail.com

Abstract
Heart failure remains a leading cause of morbidity and mortality worldwide, necessitating effective predictive models for early risk stratification. This study aims to develop and evaluate machine learning models for predicting mortality risk in heart failure patients using clinical data. A retrospective analysis was conducted on a publicly available dataset comprising 299 patients with demographic, clinical, and laboratory variables. Three supervised machine learning algorithms—Logistic Regression, Random Forest, and Support Vector Machine (SVM)—were implemented and evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC).The results demonstrated that Logistic Regression achieved the highest accuracy (0.80) and precision (0.93), while Random Forest showed the highest ROC-AUC (0.830), indicating strong discriminative ability across models. However, recall values were moderate across all models, highlighting challenges in identifying high-risk patients. Confusion matrix analysis revealed the presence of false negatives, emphasizing the need for improved sensitivity in clinical applications. Feature importance analysis identified follow-up time, serum creatinine, and ejection fraction as the most significant predictors of mortality, reflecting the combined influence of disease progression, renal dysfunction, and cardiac performance.Overall, the findings indicate that machine learning models can effectively predict mortality risk in heart failure patients using clinical data. However, further optimization and validation are required to enhance sensitivity and ensure reliable clinical application.

Keywords: *Heart failure; Mortality prediction; Machine learning; Clinical data; Logistic Regression; Feature importance*

DOI: https://doi.org/10.69980/as.v12i2.6674
Received 2 February 2026 | Accepted 30 March 2026 | Published 27 April 2026

1. Introduction

Heart failure (HF) has become a significant issue in the world, impacting millions of people, with significant morbidity, and mortality rates. It is still among the most common causes of hospitalization, especially in the elderly population, and poses a significant burden to the healthcare systems. Heart failure has a complicated and multifactorial epidemiology and etiology, with structural, functional, and metabolic heart abnormalities (Ziaeian and Fonarow, 2016). Clinical practices underscore the significance of early diagnosis, risk stratification, and proper management to enhance patient outcomes (Ponikowski et al., 2016). Irrespective of these developments, there are still high mortality rates, which underscores the necessity of more efficient predictive tools in order to identify high-risk patients (Rahimi et al., 2014).

Conventional risk-prediction techniques in heart failure have been mainly statistical models involving some pre-determined assumptions and linear correlations among variables. Nevertheless, these models might not be sufficiently able to describe the heterogeneity and complexity of the heart failure phenotypes. Recent studies have highlighted the significance of distinguishing between patient subgroups with the help of sophisticated analytical methods, including phenomapping, to enhance classification and prognosis (Shah et al., 2015). In this regard, there is an increasing desire to focus on using data-driven approaches to improve the accuracy of predictions and clinical decision-making.

Machine learning (ML) has become a potent asset in the healthcare sector, especially cardiovascular medicine, with the capacity to approximate complex and nonlinear connections between variables in large datasets (Johnson et al., 2018; Krittanawong et al., 2017). Such methods have demonstrated potential to enhance clinical outcome forecasting, such as mortality and readmission to the hospital among heart failure patients (Mortazavi et al., 2016). Moreover, the utilization of ML-based methods has become widely used in various clinical practices, showing that they can be used to facilitate decision-making and optimize patient care (Stewart et al., 2018). The idea of using machine learning and deep learning models to predict mortality risk in heart failure patients has been examined in a number of studies in recent years, with promising outcomes (Sabouri et al., 2023; Li et al., 2023).

Although such improvements have been achieved, there are still a few issues in the implementation of machine learning in clinical practice. The balance of predictive performance and interpretability is one of the main concerns. Although more complex models like deep learning algorithms can be more accurate, they tend to be less transparent and therefore are not easily applied by clinics. Simpler models, in contrast, are more interpretable but might not fully utilize the complexity of the data. Moreover, a systematic review and meta-analysis have indicated inconsistency in the performance of the models and the necessity to consider standardized assessment models in the studies of heart failure prediction (Xu et al., 2025). The need to combine model interpretability and clinical relevance to guarantee practical applicability is also highlighted in reviews of the existing machine learning applications (Kokori et al., 2025).

The other major constraint of the current literature is the use of large scale or highly specialized data groups which might not be easily accessible in every healthcare facility. Even though sophisticated artificial intelligence algorithms have been shown to predict well in a particular population (Kwon et al., 2019), their generalization is unclear. Moreover, not all studies properly discuss such clinically important evaluation measures as sensitivity and error patterns which are essential in determining the safety and reliability of predictive models in practice.

The purpose of the current work is to create and test machine learning to forecast mortality risk among patients with heart failure based on a structured clinical dataset. More specifically, the research aims to compare the performance of the Logistic Regression, Random Forest, and Support Vector machine models based on various evaluation measures, such as accuracy, precision, recall, F1-score, and ROC-AUC. Moreover, the research project seeks to measure the classification error using confusion matrix and to determine the most important predictors of mortality based on feature importance methods. This research aims to offer a wholesome and a clinically meaningful evaluation of machine learning methods in heart failure prognosis by incorporating predictive performance and model interpretability.

2. Methodology

2.1 Study Design and Data Source

This paper used a retrospective, data-driven method to build and test the machine learning models to predict mortality risk in heart failure patients. A publicly available clinical dataset, which included anonymized records of 299 patients, was analyzed (Larxel, 2019). The main outcome variable was the status of mortality at the follow-up, which was a binary variable (survival or death). The common dataset secures the reproduction and allows to perform the systematic analysis of the predictive modeling methods.

2.2 Dataset Description

The data contains 13 variables representing demographic, clinical and laboratory data that can be applied in the prognosis of heart failure. Demographic variables are age and sex whereas clinical conditions are anaemia, diabetes, high blood pressure and whether a person is a smoker. The most important physiological and lab tests are ejection fraction, serum creatinine, serum sodium, platelet count, and creatinine phosphokinase levels. It also had a follow-up variable which was the time of observation. The features were chosen based on their clinical relevance that has been established over time to determine the severity of disease and the mortality risk.

2.3 Data Preprocessing and Validation

Before the development of the model, the data was reviewed in terms of completeness and consistency. No values missing were located, and all of the categorical variables were already in binary form, which could be directly included in the analysis. Continuous variables were normalized so the features are scaled identically, and bias arising with varying magnitudes of features is reduced. The data was then separated into predictor variables and target outcome.The dataset was divided into training and testing samples in 80:20 ratio to assess the performance of the model. The models were developed using the training set and the testing set was used in an independent evaluation. Constant randomization method was used to make it reproducible. This approach gives a fair evaluation of the model on unobserved data.

2.4 Machine Learning Models

Three monitored machine learning algorithms were applied on mortality prediction. The baseline model was Logistic Regression because it is interpretable and effective when it comes to binary classification. An ensemble-based approach (Random Forest) was used to both learn nonlinear relationships between variables and to give measures of feature importance. Support Vector machine (SVM) was also employed in order to determine the best decision limits in multifaceted feature spaces. All the models were formulated in similar circumstances to achieve comparability of the results.

2.5 Model Evaluation Metrics

The evaluation metrics used to measure the model performance were various to give a holistic analysis of performance. Overall classification was measured using accuracy and precision measured the reliability of positive predictions. Recall (sensitivity) was an estimate of the capability to recognize mortality cases correctly and F1-score gave a balanced evaluation of the precision and recall. Besides that, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were employed to determine the discriminative capacity of the models at various thresholds.

2.6 Model Interpretation and Error Analysis

Error analysis and model interpretation were conducted to improve the predictive performance understanding. To assess the results of classification, a confusion matrix was applied, which allowed determining true positives, true negatives, false positives, and false negatives. The focus was particularly on false negative cases, where failure to detect high-risk patients can be of great clinical consequences.

Besides this, feature importance analysis was done with the help of the Random Forest model itself to establish the relative contribution of each predictor variable. This method also offers information on the most impactful clinical variables related to mortality and enables the interpretability of the predictive model.\

3. Results

3.1 Baseline Characteristics of the Study Population

This study involved 299 patients with heart failure. Table 1 presents the demographic and clinical features of the study population. The average age of the patients was 60.83 with a standard deviation of 11.89, which means that the group comprised mostly of middle-aged and aged people, which is also in line with the heart failure epidemiology. Mean ejection fraction was 38.08 ± 11.83%, which implied high percentage of reduced systolic functioning of heart as is characteristic of heart failure.

Laboratory values showed a mean serum creatinine of 1.39 ± 1.03 mg/dL, which indicated changes in the renal functioning of the patients. The serum sodium levels were 136.63 ± 4.41 mEq/L, which is rather stable electrolyte levels in the cohort, but minor differences could be of clinical interest. The average number of platelets was 263,358 ± 97,804 cells/µL with creatinine phosphokinase (CPK) levels significantly dispersed (581.84 ± 970.29 U/L), indicating the presence of heterogeneity in muscle or cardiac damage. The follow-up time was widely distributed, with the average time of 130.26 and a standard deviation of 77.61 days indicating that there was variation in the time of monitoring the patients and the progression of the disease.

Table 1. Baseline Characteristics of the Study Population

Variable	Mean ± SD	Min	Max
Age (years)	60.83 ± 11.89	40	95
Ejection Fraction (%)	38.08 ± 11.83	14	80
Serum Creatinine (mg/dL)	1.39 ± 1.03	0.5	9.4
Serum Sodium (mEq/L)	136.63 ± 4.41	113	148
Platelets (cells/µL)	263,358 ± 97,804	25,100	850,000
CPK (U/L)	581.84 ± 970.29	23	7861
Follow-up Time (days)	130.26 ± 77.61	4	285

3.2 Outcome Distribution

Table 2 shows the distribution of clinical outcomes. Out of 299 patients, 96 people (32.1%) died during a follow-up period, and 203 patients (67.9%) survived. This suggests a moderate level of imbalance in classes and such imbalance is typical of clinical datasets. The mortality rate of this cohort is consistent with the reported mortality in heart failure populations and it can thus be stated that the dataset is valid in predictive modeling. The importance of identifying patients at the highest risk of the adverse outcome is also highlighted by this distribution because almost one-third of the cohort had adverse outcomes.

Table 2. Distribution of Mortality Outcomes in the Study Population

Outcome	Frequency	Percentage
Survived (0)	203	67.9%

Death (1)	96	32.1%

3.3 Predictive Performance of Machine Learning Models

Three machine learning models, namely, Logistic Regression, Random Forest, and Support Vector Machine (SVM), were tested with an 8020 train-test split. Table 3 summarises the results of model performance. The overall accuracy of Logistic Regression was highest at 0.80 compared to both the Random Forest and SVM which had the accuracy of 0.75.

Logistic Regression performed better in terms of precision (0.93), showing that it is capable of accurately predicting mortality cases. The recall of the Logistic Regression was however moderate (0.56) indicating that there was a proportion of mortality cases which were not detected. Random Forest and SVM had a somewhat lower precision (0.86) and recall (0.48), which means that they were relatively less sensitive to high-risk patients. The F1-score that gives a balance between the accuracy and the recall was the best with Logistic Regression (0.70), then random forest and SVM (0.62 each).

The discriminative ability of all the models was good, with ROC-AUC values of more than 0.80. Random Forest had the best ROC-AUC value (0.830), then Logistic Regression (0.824) and SVM (0.816), which means that they are all efficient at separating survivors and non-survivors.

Table 3. Model Performance Comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC-AUC**
Logistic Regression	0.80	0.93	0.56	0.70	0.824
Random Forest	0.75	0.86	0.48	0.62	0.830
SVM	0.75	0.86	0.48	0.62	0.816

Figure 1 also demonstrates the relative performance of the models. The ROC curves indicate that the performance of all models were strong in terms of classification, with small variations in the values of AUC. In this case, although Random Forest had the highest AUC, the balance between sensitivity and precision was higher with Logistic Regression, so it is the most reliable model.

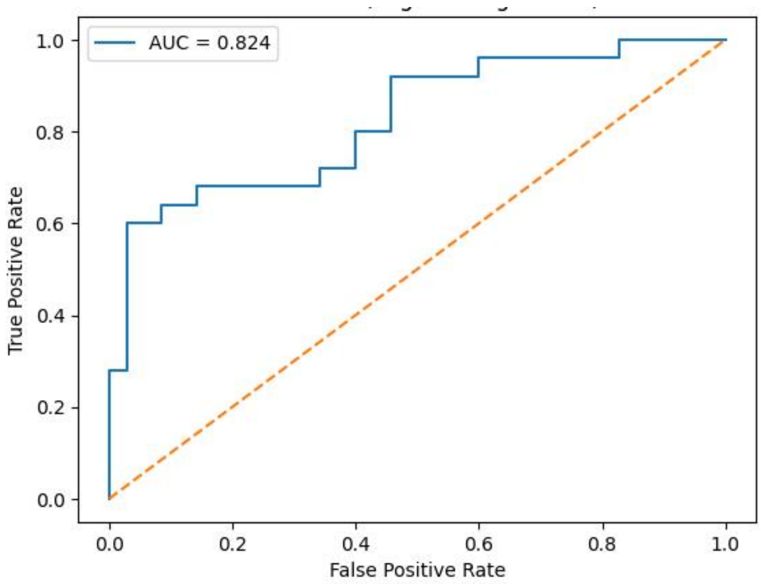

Figure 1. Receiver Operating Characteristic (ROC) Curves for Model Performance

3.4 Confusion Matrix Analysis

Figure 2 depicts a confusion matrix generated to obtain additional information on classification performance of the Logistic Regression model. The model had a true negative of 34 survivors and a true positive of 14 mortality cases. But 11 cases of mortality were mistakenly identified as survivors (false negative), and one survivor was mistakenly identified as a mortality case (false positive).

These outcomes suggest that the model was very specific as the number of false positives was very low. Nevertheless, the fact that the number of false negatives is relatively large points to the limitation of sensitivity, which is essential in the clinical setting where the inability to recognize high-risk patients can have severe repercussions. This is a significant trade-off between precision and recall that should be taken into account when using predictive models in a healthcare environment

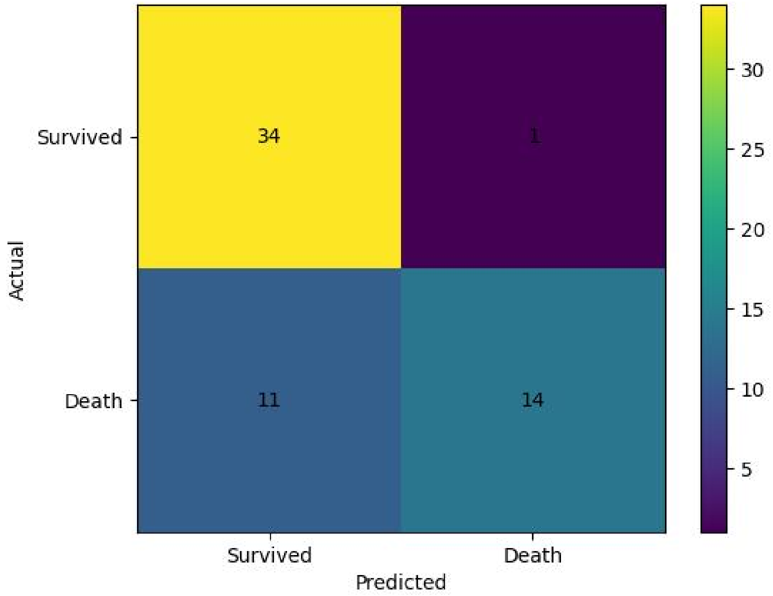

Figure 2. Confusion Matrix for Logistic Regression Model

3.5 Feature Importance Analysis

The analysis of the importance of the features was carried out with the help of the Random Forest model, to find the most significant predictors of mortality. Figure 3 shows the results. Follow-up time was found to be the most important predictor of all variables which points to the fact that the time frame of the disease progression has a critical influence on the patient outcomes.

The second most significant feature was found to be serum creatinine, which shows the influence of renal dysfunction on the risk of mortality in patients with heart failure. Ejection fraction was a predictor, too, and strengthens its accepted status as an important measure of cardiac functions. Other factors like age, platelet count and serum sodium were moderately predictive of the model whereas other factors like smoking, anaemia and high blood pressure were lowerly significant.

The results are also in line with the clinical knowledge because both cardiac and renal functions have been identified to have the severest influence on the prognosis in heart failure patients.

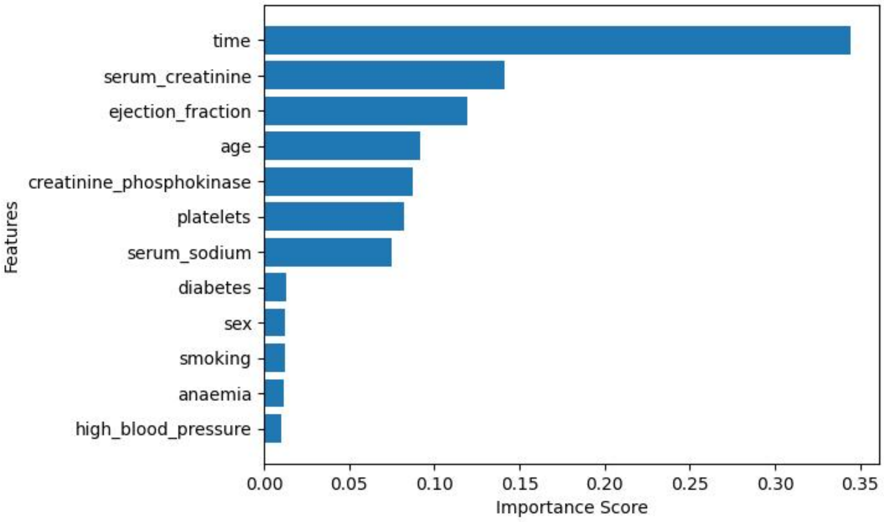

Figure 3. Feature Importance Based on Random Forest Model

The findings reveal that machine learning models can be used to accurately identify the risk of mortality in heart failure patients based on clinical data. The best overall performance was obtained with Logistic Regression which offers a good trade-off between precision and recall. All models showed good discriminative ability, but some of them had false negative, which implies that more refinements are necessary to increase sensitivity.

The analysis of feature importance also helps to prove the clinical significance of such important predictors as follow-up time, renal functioning and cardiac performance. These results not only demonstrate the potential of machine learning methods to aid in clinical decision-making, but also underline the need to properly evaluate and validate the models.

4. Discussion

The current work proves that machine learning methods can be used to forecast the risk of mortality in heart failure patients. The best overall performance of the models under consideration was that of Logistic Regression, with the accuracy of 80 per cent and the optimal balance of the precision and recall. The discriminative ability between Random Forest and ROC-AUC is a little bit higher, but the recall of the former is lower, which indicates that it is less sensitive to detecting high-risk patients. Prior research has found the same results, with traditional statistical models like the Logistic Regression outperforming or even outperforming more complicated machine learning models when trained on structured clinical data (Scrutinio et al., 2025; Sun et al., 2022). This is an indication that the model simplicity and interpretability continues to be relevant in clinical prediction problems, especially in a real-world healthcare context where transparency is critical (Beam & Kohane, 2018).

The analysis of the ROC curve in this study showed that all the models performed well in terms of classification, with an AUC value of over 0.80. Such results align with previous studies that have mentioned that machine learning methods are effective in forecasting the outcomes of heart failures (Sabouri et al., 2023; Li et al., 2023). Nevertheless, additional analysis based on the confusion matrix showed that sensitivity was quite limited, especially as indicated by the occurrence of false negative cases. This means that there was a group of patients who died but were wrongly identified as survivors. Other machine learning studies on heart failure have reported similar issues in obtaining high sensitivity and highlighted the importance of model optimization and validation (Haider et al., 2023; Ashrafi et al., 2024). In a bigger sense, these constraints are an indication of continued struggles with implementing machine learning models in clinical practice, especially in striking a balance between predictive performance and patient safety (Obermeyer and Emanuel, 2016).

The analysis of feature importance offered further information about the clinical importance of predictor variables (Topol, 2019). The most influential feature was the follow-up time which reflected the significance of the disease progression in predicting patient outcomes. This is in line with the past evidence indicating the time-dependent nature of heart failure progression and its effect on mortality (Roudini et al., 2024). It was found that serum creatinine is one of the main predictors, which indicated the importance of renal dysfunction as a predictor of heart failure. The interrelation between kidney and heart systems is proven, and high levels of creatinine are closely linked with negative outcomes (Chicco & Jurman, 2020). Likewise, ejection fraction was one of the most significant predictors, which supports its importance as a primary measure of the cardiac functioning and survival outcomes in the heart failure patients (Kwon et al., 2019).

The general literature on the use of machine learning in cardiovascular disease also supports these findings of this study. Recent reviews have highlighted the increasing importance of machine learning in enhancing the accuracy of prediction and assisting in clinical decision-making around heart failure (Kokori et al., 2025; Gungbias & Kassem, 2025). Moreover, the new methods that involve more advanced data sources and algorithms have already shown the encouraging outcomes of improving the predictive performance (Ahmadli et al., 2024). More broadly, the growing use of machine learning in medicine points to its ability to revolutionize clinical practice as it allows making decisions based on data and providing personalized care (Rajkomar et al., 2019). Even with these innovations, numerous researchers emphasize the need to strike a balance between predictive accuracy and interpretability, especially in clinical practice where transparency is crucial (Beam and Kohane, 2018).

Although the results show promise, there are a number of limitations that need to be recognized. The sample size used (n = 299) is relatively small, which can hinder the generalizability of the results and predisposes the possibility of overfitting. Also, lack of external validation limits the performance of the models among different populations. The limitations are typical of other studies employing clinical datasets and point to the necessity to validate them with larger, multi-center data (Sabouri et al., 2023; Sun et al., 2022). Moreover, the construction of powerful prediction models based on real-life clinical data has a number of methodological issues, such as the quality of the data, its heterogeneity, and bias (Goldstein et al., 2016). Some clinically relevant variables, including medication history, imaging data, and detailed comorbidities, are also not available in the dataset of this study, which can also improve the predictive performance.

The relatively low recall of all models is another significant limitation as it means that not all high-risk patients were detected. It is also a problem that was previously mentioned in other machine learning works and is a major issue with clinical prediction models (Haider et al., 2023; Ashrafi et al., 2024). Enhancement of sensitivity, without affecting the overall model performance, should hence be a major concern in future research.

Clinically, the findings of this work are indicative of the potential of machine learning models in assisting in risk stratification among patients with heart failure. Such models can help clinicians prioritize treatment and monitoring strategies by determining those at increased risk of mortality. Their significance as crucial predictors, as identified by ejection fraction and serum creatinine, further enhances their usefulness in clinical routine (Chicco & Jurman, 2020; Kwon et al., 2019). Due to the identified limitations, these models must be applied as supporting tools instead of independent decision-making systems and need to be combined with clinical knowledge.

Further development of machine learning methods, including ensemble learning and optimization schemes, might also be considered to enhance sensitivity and predictive performance. The integration of explainable methods might also be essential to enhance the transparency of the model and enable its clinical acceptance. The overall development of artificial intelligence in medical care implies that in the future, human knowledge and machine intelligence will cooperate to enhance the results of patients.

This paper shows that machine learning models can be successfully used to predict mortality risk in heart failure patients based on clinical data. Logistic Regression showed the most balanced results, and the analysis of feature importance indicated the utmost importance of cardiac and renal functioning. The results align with the available literature and can be used to validate the increasing place of machine learning in clinical decision support. Such models, however, need additional validation and refinement before they can be popularly applied in clinical practice.

References

1. Ahmadli, N., Sarsil, M. A., Mizrak, B., Karauzum, K., Shaker, A., Tulumen, E., ... & Ergen, O. (2024). Voice-driven mortality prediction in hospitalized heart failure patients: A machine learning approach enhanced with diagnostic biomarkers. *arXiv preprint arXiv:2402.13812*.
2. Ashrafi, N., Abdollahi, A., Zhang, J., & Pishgar, M. (2024). Optimizing mortality prediction for icu heart failure patients: Leveraging xgboost and advanced machine learning with the mimic-iii database. *arXiv preprint arXiv:2409.01685*.
3. Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, *20*(1), 16.
4. GUNGBIAS, H., & KASSEM, M. (2025). Death Events from Heart Failure Prediction Using Machine Learning Approach. *International Journal*, *11*(1), 1-10.
5. Haider, M. A. Y., Habibollah, H., Shukor, M. T., & Rozainy, M. H. (2023, June). Heart failure mortality prediction using machine learning. In *AIP Conference Proceedings* (Vol. 2750, No. 1, p. 040069). AIP Publishing LLC.
6. Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, *71*(23), 2668-2679.
7. Kokori, E., Patel, R., Olatunji, G., Ukoaka, B. M., Abraham, I. C., Ajekiigbe, V. O., ... & Aderinto, N. (2025). Machine learning in predicting heart failure survival: a review of current models and future prospects. *Heart failure reviews*, *30*(2), 431-442.
8. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657-2664.
9. Kwon, J. M., Kim, K. H., Jeon, K. H., Lee, S. E., Lee, H. Y., Cho, H. J., ... & Oh, B. H. (2019). Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PloS one*, *14*(7), e0219302.
10. Larxel. (2019). *Heart failure clinical data* [Data set]. Kaggle. https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data
11. Li, D., Fu, J., Zhao, J., Qin, J., & Zhang, L. (2023). A deep learning system for heart failure mortality prediction. *PLoS One*, *18*(2), e0276835.
12. Mortazavi, B. J., Downing, N. S., Bucholz, E. M., Dharmarajan, K., Manhapra, A., Li, S. X., ... & Krumholz, H. M. (2016). Analysis of machine learning techniques for heart failure readmissions. *Circulation: Cardiovascular Quality and Outcomes*, *9*(6), 629-640.
13. Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J. G., Coats, A. J., ... & Van Der Meer, P. (2016). 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Polish Heart Journal (Kardiologia Polska)*, *74*(10), 1037-1147.
14. Rahimi, K., Bennett, D., Conrad, N., Williams, T. M., Basu, J., Dwight, J., ... & MacMahon, S. (2014). Risk prediction in patients with heart failure: a systematic review and analysis. *JACC: Heart Failure*, *2*(5), 440-446.
15. Roudini, B., Khajehpiri, B., Moghaddam, H. A., & Forouzanfar, M. (2024). Machine learning predicts long-term mortality after acute myocardial infarction using systolic time intervals and routinely collected clinical data. *Intelligent Medicine*, *4*(03), 170-176.

16. Sabouri, M., Rajabi, A. B., Hajianfar, G., Gharibi, O., Mohebi, M., Avval, A. H., ... & Shiri, I. (2023). Machine learning based readmission and mortality prediction in heart failure patients. *Scientific Reports*, *13*(1), 18671.
17. Scrutinio, D., Amitrano, F., Guida, P., Coccia, A., Pagano, G., D'addio, G., & Passantino, A. (2025). Prediction of mortality in heart failure by machine learning. Comparison with statistical modeling. *European Journal of Internal Medicine*, *133*, 106-112.
18. Shah, S. J., Katz, D. H., Selvaraj, S., Burke, M. A., Yancy, C. W., Gheorghiade, M., ... & Deo, R. C. (2015). Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*, *131*(3), 269-279.
19. Stewart, J., Sprivulis, P., & Dwivedi, G. (2018). Artificial intelligence and machine learning in emergency medicine. *Emergency Medicine Australasia*, *30*(6), 870-874.
20. Sun, Z., Dong, W., Shi, H., Ma, H., Cheng, L., & Huang, Z. (2022). Comparing machine learning models and statistical models for predicting heart failure events: a systematic review and meta-analysis. *Frontiers in Cardiovascular Medicine*, *9*, 812276.
21. Xu, Z., Hu, Y., Shao, X., Shi, T., Yang, J., Wan, Q., & Liu, Y. (2025). The efficacy of machine learning models for predicting the prognosis of heart failure: a systematic review and meta-analysis. *Cardiology*, *150*(1), 79-97.
22. Ziaeian, B., & Fonarow, G. C. (2016). Epidemiology and aetiology of heart failure. *Nature Reviews Cardiology*, *13*(6), 368-378.
23. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358.
24. Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. (2016). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association: JAMIA*, *24*(1), 198.
25. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, *375*(13), 1216.
26. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, *319*(13), 1317-1318.
27. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56.