ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: https://dergipark.org.tr/ij3dptdi

COMPARISON OF MACHINE LEARNING MODELS IN HEART FAILURE PREDICTION AND THEIR INTEGRATION INTO CLINICAL DECISION SUPPORT SYSTEMS

Yazarlar (Authors): Mustafa Çakır*

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): *Çakır M., "Comparison of Machine Learning Models In Heart Failure Prediction and Their Integration Into Clinical Decision Support Systems" Int. J. of 3D Printing Tech. Dig. Ind., 9(2): 272-282, (2025).*

DOI: 10.46519/ij3dptdi.1724620

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): https://dergipark.org.tr/en/pub/ij3dptdi/archive

COMPARISON OF MACHINE LEARNING MODELS IN HEART FAILURE PREDICTION AND THEIR INTEGRATION INTO CLINICAL DECISION SUPPORT SYSTEMS

Mustafa Çakır[a]*

[a]Iskenderun Technical University, Iskenderun Vocational School of Higher Education, Electronics and Automation Department, TÜRKİYE

* *Corresponding Author: mustafa.cakir@iste.edu.tr*

(**Received:** 21.06.25; **Revised:** 23.07.25; **Accepted:** 08.08.25)

ABSTRACT

Heart failure remains a leading cause of morbidity and mortality worldwide, necessitating advanced tools for early risk prediction. This study presents an interactive, machine learning-driven web application designed to predict heart failure outcomes using clinical data. Leveraging the heart failure clinical records dataset (n=299), the application integrates a comprehensive suite of fifteen diverse predictive models, encompassing traditional/statistical-based algorithms, instance-based and probabilistic methods, various tree-based and ensemble techniques, and neural networks within an intuitive Shiny framework. Key features include exploratory data analysis (correlation matrices, feature importance), model training, and real-time risk prediction with customizable patient parameters. The system employs stratified cross-validation (10-fold) for robust evaluation and achieves impressive performance, with top-performing models exhibiting test set Area Under Curve values exceeding 0.85, alongside high scores in accuracy, sensitivity, specificity, and F1-score. By combining clinical variables such as ejection fraction, serum creatinine, and follow-up time, the tool demonstrates how interactive machine learning platforms can enhance clinical decision-making. The open-source R-Shiny implementation provides immediate visual feedback, model interpretability features, and a template for extending predictive analytics to other medical domains. This work bridges the gap between statistical modeling and clinical application, offering both a prognostic tool and an educational resource for data-driven cardiology.

Keywords: Heart Failure Prediction, Machine Learning, Clinical Decision Support, R-Shiny.

1. INTRODUCTION

Heart-failure (HF) is a chronic and progressive cardiovascular disease where the heart is unable to pump enough blood to meet the body's needs [1]. Affecting millions globally, HF represents a significant health challenge associated with high morbidity and mortality rates, severely diminishing patients' quality of life and imposing a substantial burden on healthcare systems [2]. Early diagnosis, accurate prognosis prediction, and the determination of effective treatment strategies are critically important in managing HF patients. However, the complex pathophysiology of HF and the variability of the disease among individuals can make precise diagnosis and prognosis challenging using traditional clinical methods.

In-recent years, the increasing availability of large datasets in medicine and advancements in computing technologies have opened up significant opportunities for applying machine learning (ML) techniques in healthcare. ML algorithms can analyse complex medical data, such as clinical findings, laboratory results, and imaging data, to uncover hidden patterns and relationships, providing valuable insights for tasks like disease diagnosis, predicting patient prognosis, and identifying risk factors that may influence disease progression. Particularly in the field of HF, ML models hold the potential to analyze patient data to predict the presence of the disease, forecast patient survival probability, and pinpoint risk factors.

A review of the literature clearly indicates a significant increase in the number of scientific publications addressing both ML and HF over the years. As illustrated in Figure 1, research interest in this area has grown considerably, especially in recent years. This trend vividly demonstrates the increasing acceptance and potential of ML techniques in HF research.

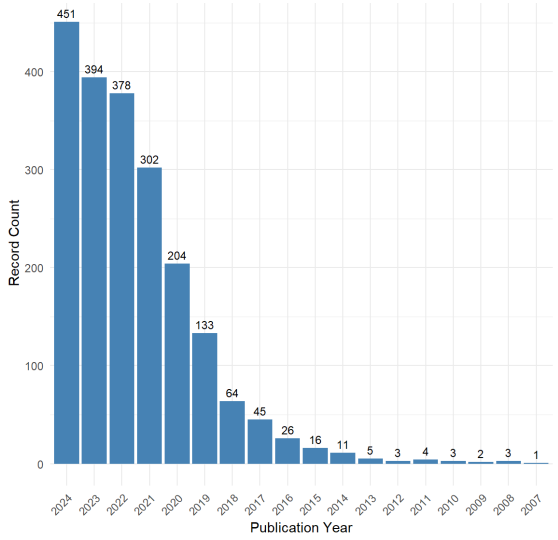


Figure 1. Trend of publications in HF and ML research

The literature frequently reveals the use of various ML algorithms for different prediction and classification tasks related to HF. These algorithms play a significant role in understanding, diagnosing, and predicting the prognosis of the disease.

A substantial portion of ML studies related to HF has focused on predicting the presence of the disease or forecasting patients' survival probability. For instance, [3] predicted the risk of death in HF patients using RF and Binary Particle Swarm Intelligence methods. [4], on the other hand, analysed the performance of Ensemble ML methods for predicting the survival of patients diagnosed with HF. Similarly, [5] proposed an ML-based approach to detect the survival status of HF patients. [6] also conducted an application using classification-based ML algorithms for predicting the survival of HF patients. Among more recent work, [7] developed time-adaptive ML models to predict the severity of HF with Reduced Ejection Fraction.

Researchers have employed a variety of ML algorithms in HF studies, including Linear Regression (LR) [4], [8–10], Support Vector Machine (SVM) [6, 7, 11], Decision Tree (DT) [12], Random Forest (RF) [13], Artificial Neural Network (ANN) [14-15], and Naïve Bayes (NB) [16]. The performance of these algorithms is typically evaluated using standard metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) [7], [10], [17–19]. Some studies have aimed to compare the performance of different algorithms on the same dataset. For example, [9], [20] conducted a comparison of different ML classification algorithms for cardiovascular disease prediction. [10] examined the performance comparison of various ML algorithms in the early diagnosis of HF. [21] presented a comprehensive study evaluating the performance of different types of ML methods categorized into Tree, Meta, and Function categories for HF prediction. The use of web-based analytical tools [22] and ML applications shows significant potential not only in heart failure prediction but also in other health domains, such as estimating COVID-19 mortality [23] in Turkey.

While the datasets used vary, many studies have preferred publicly available datasets such as the Kaggle or University of California Irvine ML Repository. Nevertheless, some research has been conducted on specific clinical datasets.

Studies within the existing literature clearly demonstrate the strong potential of ML in the field of HF. However, the relative performance of different algorithms on specific datasets and their potential for clinical application remain an active area of research.

This study aims to contribute to the field by deeply analyzing the performance of various ML models, including or in comparison with previously used algorithms, on a specific dataset. In particular, comparing the strengths and weaknesses of different model types can help in making more informed choices for the development of future clinical decision support systems. Therefore, specifically aims to compare the performance of different ML models in predicting heart failure outcomes and to evaluate their potential for integration into clinical decision support systems. Utilizing a relevant and up-to-date dataset, we will train,

test, and compare various ML algorithms based on established performance metrics. The findings from this research are expected to contribute to the development of ML-based decision support systems [24], thereby facilitating more effective management of HF patients.

2. MATERIAL AND METHOD

In this study, ML models were used to predict the occurrence of deaths in patients with HF. Details of the dataset used, algorithms applied, and the developed R-Shiny platform are described below.

2.1. Dataset

The Heart Failure Clinical Records Dataset comprises clinical information from 299 patients with HF. It contains 12 clinical features and the target variable, "DEATH_EVENT". The features included in the dataset are; age, anaemia status (anaemia), creatinine phosphokinase level (creatinine_phosphokinase), diabetes status (diabetes), left ventricular ejection fraction (ejection_fraction), high blood pressure status (high_blood_pressure), platelet count (platelets), serum creatinine level (serum_creatinine), serum sodium level (serum_sodium), sex, smoking status (smoking), and follow-up time (time). The target variable, DEATH_EVENT, is a binary variable (yes/no) indicating whether the patient passed away during the follow-up period.

Prior to model development, a comprehensive exploratory data analysis was conducted to understand the distribution of each feature and the relationships between them. As part of this analysis, the pairwise Pearson correlation coefficients among all independent clinical features were calculated and are presented as a heatmap in Figure 2. This visualization allowed for the identification of potential linear relationships between variables, which is crucial for understanding the underlying structure of the dataset and for addressing issues such as multicollinearity in subsequent predictive modeling. For instance, the Figure highlights a notable moderate positive correlation between "sex" and "smoking" (r=0.45), suggesting a potential gender-specific association with smoking status within this patient cohort. Other weaker correlations, such as the slight positive association between 'age' and "serum_creatinine" (r=0.16) or the weak negative correlation between "age" and "time" (r=-0.22), were also observed. While these correlations do not imply causation, they provide valuable insights into the interdependencies of the clinical parameters and informed the subsequent stages of feature selection and model building, ensuring robustness and interpretability of the derived models.

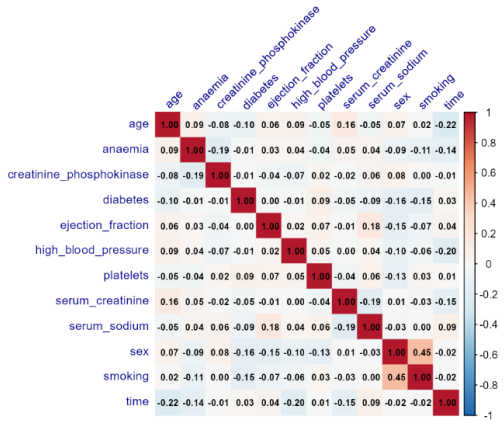


Figure 2. Feature correlation matrix

As illustrated in Figure 3, the RF feature importance analysis highlights time as the most influential predictor of death events in HF patients, with an importance score that markedly exceeds all other variables. This finding suggests that the duration of patient follow-up is a critical determinant of mortality risk, likely reflecting survival time as a proxy for event occurrence. Serum creatinine and ejection fraction emerge as the second and third most important predictors, reinforcing their established roles as indicators of renal dysfunction and cardiac performance, respectively, both of which are key prognostic factors in HF. Variables such as age, platelets, creatinine phosphokinase, and serum sodium show moderate predictive value, implying their supplementary relevance. Conversely, features including high blood pressure, anaemia, smoking, sex, and diabetes contribute minimally, indicating limited independent prognostic utility after controlling for more dominant clinical factors. Thus, the Figure provides a data-driven hierarchy of risk factors that corroborates clinical expectations and enhances our understanding of variable importance in HF prognosis.

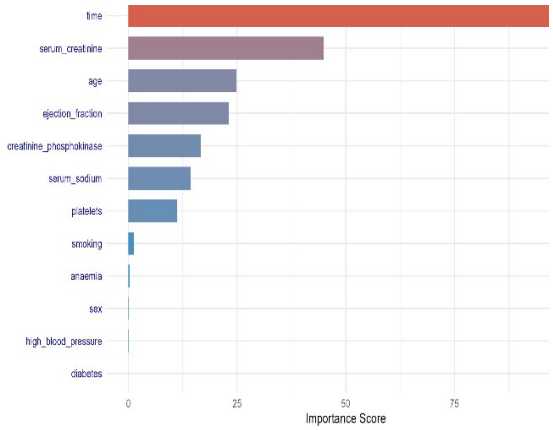


Figure 3. RF feature importance graph

2.2. Machine Learning Algorithms

In this study, various ML algorithms were compared to predict death events due to HF. The selected algorithms represent different learning approaches. The 15 different ML algorithms evaluated are listed below:

LR was introduced by David Cox in 1958 as a method to model binary outcomes using the logistic function, building on earlier work by Joseph Berkson in the 1940s who coined the term "logit" [25]. It models the log-odds of a binary response as a linear function of predictor variables, and it remains foundational in epidemiology and biomedical sciences.

Linear Discriminant Analysis (LDA) was developed by Ronald A. Fisher in 1936 as a method to find a linear combination of features that best separates two or more classes. It is widely used in pattern recognition and statistics for dimensionality reduction and classification [26].

QDA is an extension of LDA that emerged later as statisticians generalized discriminant analysis to relax the assumption of equal covariance matrices [27].

K-Nearest Neighbor (KNN) was first described by Evelyn Fix and Joseph Hodges in 1951 as a non-parametric method for pattern classification [28]. It gained prominence after Thomas Cover and Peter Hart formalized its statistical properties in 1967 [29].

NB classifiers stem from Bayes' Theorem proposed by Thomas Bayes in the 18th century. The "naive" assumption of conditional independence was formalized and applied in pattern recognition and information retrieval starting in the 1950s, with early notable applications in text classification by Maron in 1961 [30].

DT gained traction through the Classification and Regression Trees (CART) framework developed by Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone [31]. The CART methodology popularized recursive partitioning for classification and regression tasks.

RF was introduced by Leo Breiman in 2001 as an ensemble learning method that aggregates predictions of multiple decision trees to improve accuracy and reduce overfitting [32].

BT Bagging, short for Bootstrap Aggregating, was introduced by Leo Breiman in 1996 as a method that enhances the stability and predictive accuracy of ML algorithms by reducing variance through the aggregation of multiple bootstrapped models. This ensemble technique is particularly effective when applied to high-variance models such as decision trees because it combines the predictions of several models trained on different subsets of the data to improve overall performance and mitigate overfitting [33].

Fast Random Forest (FRF) is a fast implementation of RF introduced by Marvin N. Wright and Andreas Ziegler in 2017, optimized for high-dimensional data and genome-wide association studies [34].

Gradient Boosting Machine (GBM), the concept of boosting was introduced by [35] and further advanced by [36]. The specific formulation of gradient boosting was developed by [37], which generalized boosting algorithms using gradient descent techniques.

SVM was developed by Vladimir Vapnik and Alexey Chervonenkis in the 1960s, and popularized in the 1990s through Vapnik's extensive work [38]. The introduction of kernel functions, including the Radial Basis Function, extended its applicability to non-linear problems.

ANN, the foundational idea of artificial neural networks was introduced by Warren McCulloch and Walter Pitts in 1943. The modern, multilayer perceptron model and backpropagation algorithm were popularized by [39].

XGBoost, short for "Extreme Gradient Boosting" was introduced by Tianqi Chen and Carlos Guestrin in 2016 as a scalable and regularized gradient boosting framework optimized for efficiency and predictive performance [40].

GLMNET was introduced by Hui Zou and Trevor Hastie in 2005 as a regularization technique that combines the strengths of L1 (Lasso) and L2 (Ridge) penalties to handle correlated predictors [41].

Table 1. Evaluation metrics

Metric	*Definition*	*Formula*
Accuracy	The ratio of correctly classified instances (both positive and negative) to the total number of instances.	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (Recall)	The proportion of actual positive cases (patients who passed away) that were correctly identified by the model.	$\frac{TP}{TP+FN}$
Specificity	The proportion of actual negative cases (patients who did not pass away) that were correctly identified by the model.	$\frac{TN}{TN+FP}$
Precision	The proportion of predicted positive cases that are actually positive.	$\frac{TP}{TP+FP}$
F1 Score	The harmonic mean of Precision and Sensitivity, serving as a balanced measure of model performance, especially in imbalanced datasets.	$2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$

2.3. R-Shiny Application Description

The structure and workflow of the R-Shiny-based application are illustrated in Figure 4. As depicted, the system initiates with dataset loading and preprocessing, and then branches into multiple functional tabs, each dedicated to a specific aspect of the machine learning pipeline, including data exploration, model training, evaluation, prediction, and interpretability.

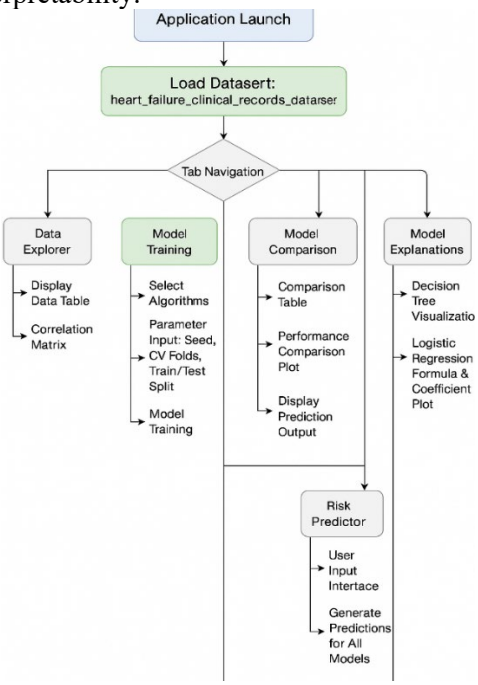


Figure 4. Workflow of R-Shiny based application

3. RESULTS AND DISCUSSION

For model training and performance evaluation, the dataset was split into two parts: 90% for training and 10% for testing. This split ensures that the models are tested on an independent dataset they have not seen during training, allowing for a more accurate estimation of their generalization capabilities. A fixed random seed value was used to reduce variations in results caused by randomness in data splitting and model training.

During the model training phase, 10-fold cross-validation was applied to estimate model performance more reliably and to optimize the algorithms' tuning parameters. In cross-validation, the training portion of the dataset was divided into 10 equal parts; in each iteration, one part was held out for validation while the remaining nine parts were used for training. This process was repeated 10 times, and the average performance was recorded. The optimal tuning parameters for the models were determined by selecting the combination that yielded the highest Area Under the ROC Curve (AUC) value from the cross-validation results.

The performance of the trained ML models was evaluated on the independent test dataset using various metrics. These metrics provide information about the models' prediction accuracy and classification abilities:

Upon analyzing the experimental results, it is evident that the ML models exhibited varying levels of success in predicting the HF death event. The evaluation conducted on the independent test dataset (Table 2) revealed that several models demonstrated superior overall performance. Specifically, GLM, LDA, GBM, SVM Linear, and GLMNET achieved the highest overall performance metrics, with Accuracy (96.6%) and F1_Score (97.6%) reaching and, respectively.

Table 2. Performance Evaluation Of Different Ml Models

Algorithm	Accuracy	Sensitivity	Specificity	F1_score
LR	0.966	1	0.889	0.976
LDA	0.966	1	0.889	0.976
GLMNET	0.966	1	0.889	0.976
GBM	0.966	1	0.889	0.976
SVM Linear	0.966	1	0.889	0.976
XGBoost	0.931	1	0.778	0.952
RF	0.931	1	0.778	0.952
FRF	0.931	1	0.778	0.952
SVM Radial	0.897	0.95	0.778	0.927
BT	0.897	0.95	0.778	0.927
DT	0.862	0.85	0.889	0.895
NB	0.828	0.95	0.556	0.884
NN	0.828	0.75	1	0.857
QDA	0.793	0.95	0.444	0.864
KNN	0.586	0.8	0.111	0.727

A particularly significant finding for these top-performing models is their Sensitivity value of 1.00 on the test set, indicating that they correctly identified

all the death cases in the test dataset. This highlights their substantial potential for identifying high-risk patients in a clinical setting. Other robust models, such as RF, FRF, and XGBoost, also performed strongly, with Accuracy (93.1%) and F1_Score 95.2%. These models also achieved a Sensitivity of 1.00, though their Specificity (77.8%) was slightly lower than that of the leading group.

The interpretable nature of the LR model, as shown by its equation (Equation 1) and coefficient plot (Figure 5), provided clear insights into the direction and magnitude of the relationship between individual clinical factors and the log-odds of death, largely aligning with clinical understanding of HF prognostic indicators.

The LR Coefficients plot:

$$Logit(p) = 9.23 + 0.041 * age - 0.05 * anaemia + 0.206 * diabetes - 0.077 * ejection\_fraction - 0.003 * high\_blood\_pressure + 0.667 * serum\_creatinine - 0.058 * serum\_sodium + -0.485 * sex - 0.041 * smoking - 0.019 * time \quad (1)$$

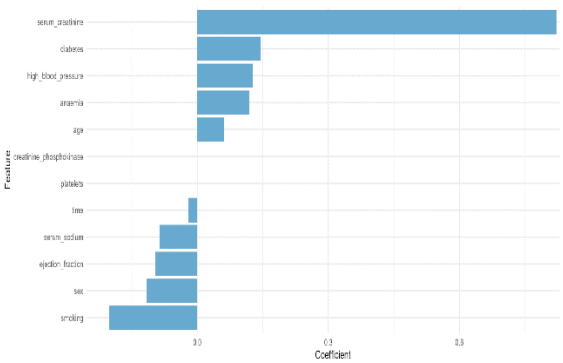


Figure 5. LR coefficient plot

This plot scientifically visualizes the estimated coefficients of the LR model, quantifying the linear association between each clinical feature and the log-odds of the HF death event, holding other variables constant. Features with positive coefficients, represented by bars extending to the right (notably serum creatinine and diabetes), indicate an increased log-odds, and thus an increased probability, of mortality as their values increase. Conversely, features with negative coefficients, shown by bars extending to the left (such as sex and ejection fraction), suggest a decreased log-odds of death with increasing values or different categories, highlighting their protective or inverse relationship with mortality risk in this model. Features with coefficients close to zero, indicated by bars near the central axis, demonstrate a minimal linear impact on the log-odds of death when considered alongside other predictors in the model. This visual representation is crucial for interpreting the model's internal structure, revealing which clinical factors are most strongly associated with the outcome and in what direction, thereby contributing to the scientific understanding of feature-outcome relationships within this specific linear framework.

Figure 6 illustrates the constructed DT model, which clearly identifies key clinical features and their respective thresholds in predicting patient outcome. The tree's root node initiates with "time", indicating that patients with a follow-up time less than 74 days (time<74) enter a distinct risk pathway. Within this subgroup, "ejection_fraction" emerges as the next crucial predictor: patients with "ejection_fraction>=33" exhibit a remarkably low mortality rate of 6% (representing 54% of the total dataset), suggesting this as the lowest-risk cohort. Conversely, for those with "ejection_fraction<33", the "serum_creatinine" level becomes decisive; patients with "serum_creatinine>=1.4" face a high mortality rate of 65% (comprising 9% of the dataset), identifying this as a particularly high-risk group characterized by shorter follow-up, reduced ejection fraction, and elevated serum creatinine. Intriguingly, patients with a follow-up time of 74 days or more (time>=74), constituting 26% of the dataset, show a high mortality rate of 81%. This highly interpretable structure allows clinicians to rapidly identify specific patient subgroups based on these key clinical parameters and their thresholds, facilitating more informed prognostic assessments.

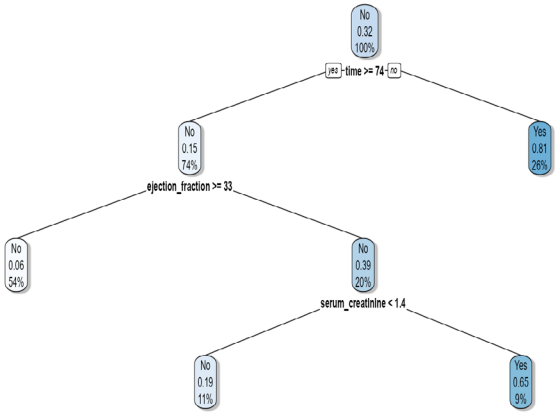


Figure 6. Visual representation of the DT model for predicting death event

Regarding computational efficiency, the training duration analysis (Figure 7) showed that ensemble methods like XGBoost, Ranger, and RF generally required longer training times compared to simpler models such as LR, LDA, and KNN. This suggests a trade-off between model complexity/training time and performance, which is a crucial consideration for practical deployment.

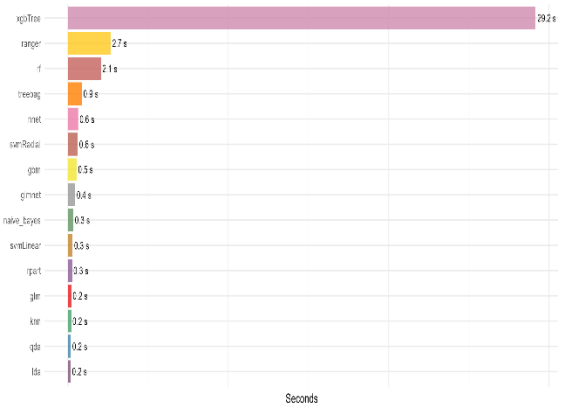
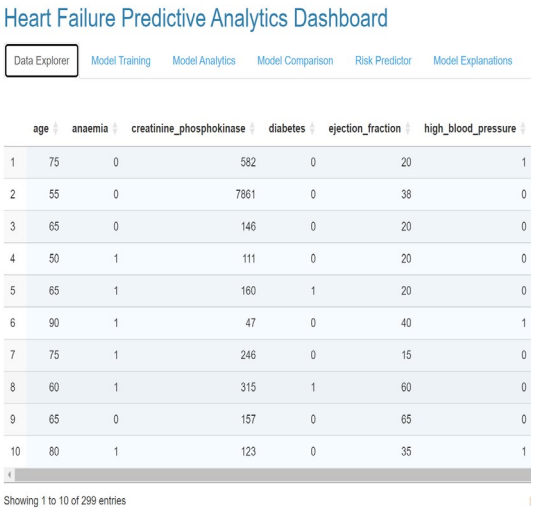


Figure 7. Model training duration

The developed R-Shiny application serves as a practical tool to leverage these trained models in a clinical context. Figure 8 provides an overview of the R-Shiny application's user interface, specifically showcasing the "Data Explorer" tab which facilitates initial data inspection and preliminary feature review. Beyond data exploration, the application integrates various functionalities such as model training, detailed model analytics, and model comparison. The "Risk Predictor" functionality, for instance, allows users to input specific patient clinical data and obtain instantaneous

risk predictions from multiple models. This feature demonstrates the potential of integrating these ML models into a clinical decision support system, offering healthcare professionals an additional resource to aid in risk assessment and decision-making by providing predictions based on diverse algorithmic perspectives.



	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
1	75	0	582	0	20	1
2	55	0	7861	0	38	0
3	65	0	146	0	20	0
4	50	1	111	0	20	0
5	65	1	160	1	20	0
6	90	1	47	0	40	1
7	75	1	246	0	15	0
8	60	1	315	1	60	0
9	65	0	157	0	65	0
10	80	1	123	0	35	1

Figure 8. R-Shiny application user interface

4. CONCLUSION

This study successfully demonstrated the robust potential of various ML models in predicting heart failure outcomes, particularly death events, by leveraging the Heart Failure Clinical Records Dataset. My interactive, ML-driven web application, developed within the intuitive R-Shiny framework, served as a comprehensive platform for evaluating fifteen diverse predictive algorithms, encompassing traditional/statistical-based methods, instance-based and probabilistic methods, various tree-based and ensemble techniques, and neural networks.

The experimental results revealed that several models, including GLM, LDA, GBM, SVM Linear, and GLMNET, exhibited superior overall performance on the independent test dataset, achieving an Accuracy of 96.6% and an F1_Score of 97.6%. A particularly significant finding was the Sensitivity of 1.00 for these top-performing models, indicating their exceptional ability to correctly identify all death cases within the test set. This highlights their substantial promise for pinpointing high-risk patients in clinical environments. Other robust models like RF, Ranger, and XGBoost also showed strong performance with 93.1% Accuracy and 95.2% F1_Score, alongside perfect Sensitivity.

The application's key features, such as exploratory data analysis (including correlation matrices and feature importance), model training, and real-time risk prediction, illustrate the practical utility of integrating these models into clinical decision support systems. The interpretability provided by models like LR and DT offers valuable insights into the relationships between clinical factors and mortality risk, corroborating existing medical understanding.

In conclusion, this work bridges the gap between statistical modeling and clinical application, providing both a prognostic tool and an educational resource for data-driven cardiology. The findings from this research are expected to contribute to the development of ML-based decision support systems, thereby facilitating more effective management of HF patients.

For future studies on this research topic, several key areas warrant further investigation. From a quantitative research perspective, future work should involve validating these models on larger and more diverse datasets, ideally from multiple institutions or different geographic regions, to enhance their generalizability and external validity. This expansion would provide more robust evidence of model performance across varied patient populations. Conducting prospective studies in actual clinical settings will also be crucial to evaluate the real-world impact and effectiveness of this interactive ML platform in improving patient outcomes and aiding healthcare professionals in their decision-making processes, which involves integrating the tool into clinical workflows and assessing its utility in a live environment. Additionally, quantitative research could focus on the economic impact of using such decision support systems, evaluating potential cost savings related to early prediction and optimized treatment strategies, and exploring the integration of real-time data from wearable health devices or continuous monitoring systems to enhance predictive accuracy and enable more dynamic risk assessments. Conversely, qualitative research is also vital. This includes understanding the perspectives of healthcare professionals regarding the usability, trustworthiness, and overall acceptability of

ML-based clinical decision support tools through interviews and focus groups, which could explore barriers and facilitators to adoption. Further qualitative research could delve into the ethical considerations surrounding the deployment of AI in clinical settings, particularly concerning data privacy, algorithmic bias, and accountability in decision-making. Finally, understanding the specific training and educational requirements for clinicians to effectively utilize and interpret ML model outputs is another crucial qualitative research area. By addressing these quantitative and qualitative research avenues, the field can further advance the integration of ML into real-world clinical practice for heart failure management.

ACKNOWLEDGES

This work was presented at the 7th International Mediterranean Scientific Research and Innovation Congress. The author expresses their gratitude to the UCI Machine Learning Repository for providing access to the Heart Failure Clinical Records Dataset, which was fundamental to this study. The author also acknowledges the open-source community and developers of the R-Shiny framework, whose accessible and powerful tools enabled the development of the interactive analytical application central to this research. The code used in this study has been shared as open-source in the GitHub repository (https://github.com/ckrmustafa/HeartDiseasePrediction).

REFERENCES

1. McDonagh, T. A. et al., “2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure,” Eur. Heart J., Vol. 42, Issue. 36, Pages. 3599–3726, 2021.

2. L. S. Ranard, S. A. Parikh, and A. J. Kirtane, “COVID-19–Specific Strategies for the Treatment of ST-Segment Elevation Myocardial Infarction in China,” J. Am. Coll. Cardiol., Vol. 76, Issue. 11, Pages. 1325–1327, 2020.

3. Salman, O. K. M. and Aksoy, B. “Rasgele Orman Ve İkili Parçacik Sürü Zekâsi Yöntemiyle Kalp Yetmezliği Hastalığındaki Ölüm Riskinin Tahminlenmesi,” Int. J. 3D Print. Technol. Digit. Ind., Vol. 6, Issue. 3, Pages. 416–428, 2022.

4. Yapıcı, İ. Ş., Arslan, R. U., and Erkaymaz, O. “Kalp Yetmezliği Tanılı Hastaların Hayatta Kalma Tahmininde Topluluk Makine Öğrenme Yöntemlerinin Performans Analizi,” Karaelmas Fen ve Mühendislik Derg., Vol. 14, Issue. 1, Pages. 59–69, 2024.

5. Erdaş, B., and Ölçer, D., “Kalp Yetmezliği Hastalarının Hayatta Kalma Tespiti İçin Makine Öğrenmesi Tabanlı Bir Yaklaşım,” Pages. 16–19, 2020.

6. Keser, S. B., and Keskin, K., “Kalp Yetmezliği Hastalarının Sağ Kalım Tahmini: Sınıflandırmaya Dayalı Makine Öğrenmesi Algoritmalarının Bir Uygulaması,” Afyon Kocatepe Univ. J. Sci. Eng., Vol. 23, Issue. 2, Pages. 362–369, 2023.

7. Winger, T., Ozdemir, C., Narasimhan, S. L. and Srivastava, J., “Time-Adaptive Machine Learning Models for Predicting the Severity of Heart Failure with Reduced Ejection Fraction,” Diagnostics, Vol. 15, Issue. 6, Pages. 1–11, 2025.

8. Aydemir, M., Çakir, M., Oral, O., and Yilmaz, M., “Diagnosis of Cushing’s syndrome with generalized linear model and development of mobile application,” Medicine (Baltimore)., Vol. 104, Issue. 25, Pages. e42910, 2025.

9. Kaba G., and Kalkan, S. B., “Kardiyovasküler Hastalık Tahmininde Makine Öğrenmesi Sınıflandırma Algoritmalarının Karşılaştırılması,” İstanbul Ticaret Üniversitesi Fen Bilim. Derg., Vol. 21, Issue. 42, Pages. 183–193, 2022.

10. Gürgen, G., and Serttaş, S., “Kalp Yetmezliği Hastalığının Erken Teşhisinde Makine Öğrenimi Algoritmalarının Performans Karşılaştırması,” Euroasia J. Math. Eng. Nat. Med. Sci., Vol. 10, Pages. 165–174, 2023.

11. Nasution, N., Nasution, F. B., and Hasan, M. A., “Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset,” IT J. Res. Dev., Vol. 9, Issue. 2, Pages. 140–150, 2025.

12. Yang, M., Zhang, X., Fang, W., Zhang, Y., and Fan, X., “Trajectories and Predictors of Frailty in Patients with Heart Failure: A Longitudinal Study,” J. Clin. Nurs., Vol. 0, Pages. 1–14, 2025.

13. Ribeiro, E. G. et al., “Effect of Telemedicine Interventions on Heart Failure Hospitalizations: A Randomized Trial,” J. Am. Heart Assoc., Vol. 14, Issue. 6, p. e036241, 2025.

14. Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., and Salar, S. A. A., "Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan," Springer, Vol. 25, Issue 1, 2025.

15. Mondal, S., Maity, R., Reports, A. N.-S., "An efficient artificial neural network-based optimization techniques for the early prediction of coronary heart disease: comprehensive analysis," nature.com, Vol. 15, Issue. 4827, Pages. 1–24, 2025.

16. Escobar, D. A. C., and Rivera, M. M., "Prediction of Heart Disease Using a Pattern Recognition Approach with Feature Selection and Naïve Bayesian Classifier," Springer, Pages. 347–363, 2025.

17. Chitta, S., Chandu, S., Chaitanya, K., Katha, Sundar Junapudi, S., Kumar Yata, V. and Junapudi, S., "Clinical and demographic predictors of heart failure outcomes: A machine learning perspective," Eurasian J. Med. Oncol., Vol. 9, Issue. 1, Page. 133, Jan. 2025.

18. Chulde-Fernández, B., et al., "Classification of Heart Failure Using Machine Learning: A Comparative Study," MDPI:Life, Vol. 15, Issue. 3, Pages. 1–18, 2025.

19. Khalid Hussain, M., Wani, S., and Abubakar, A., "Examining Mortality Risk Prediction Using Machine Learning in Heart Failure Patients," Int. J. Perceptive Cogn. Comput., Vol. 11, Issue. 1, Pages. 81–87, Jan. 2025.

20. Akbulut, B., Çakır, M., Görkem Sarıkaya, M., Oral, O., Yılmaz, M., and Aykal, G., "Artificial intelligence to predict biomarkers for new-onset atrial fibrillation after coronary artery bypass grafting," Turkish J. Thorac. Cardiovasc. Surg., Vol. 33, Issue. 2, Pages. 144–153, Apr. 2025.

21. Atacak, İ., "Kalp Yetmezliği Tahmininin Kategorik Olarak Farklı Tip Makine Öğrenmesi Yöntemleri ile Uygulanmasına Yönelik Bir Değerlendirme Çalışması," EMO Bilim. Dergi, Vol. 14, Issue. 1, Pages. 73–95, 2024.

22. Yigit, T., Isik, A. H., and Ince, M., "Web-based learning object selection software using analytical hierarchy process," IET Softw., Vol. 8, Issue. 4, Pages. 174–183, 2014.

23. Turan, T., Turan, G., and Köse, U., "Uyarlamalı Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi ve Yapay Sinir Ağları ile Türkiye'deki COVID-19 Vefat Sayısının Tahmin Edilmesi," Bilişim Teknol. Derg. Vol. 15, Issue. 2, Pages. 97–105, 2022.

24. Almazroi, A. A., Aldhahri, E. A., Bashir, S., and Ashfaq, S., "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," IEEE Access, Vol. 11, Issue June, Pages. 61646–61659, 2023.

25. Cox, D. R., "The Regression Analysis of Binary Sequences," J. R. Stat. Soc. Ser. B Stat. Methodol., Vol. 20, Issue. 2, Pages. 215–232, 1958.

26. Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems," Ann. Eugen., Vol. 7, Issue. 2, Pages. 179–188, 1936.

27. Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning. New York, NY: Springer New York, 2009.

28. Fix E., and Hodges, J. L., "Discriminatory analysis. Nonparametric discrimination: Small sample performance," 1951.

29. Cover T., and Hart, P., "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, Vol. 13, Issue. 1, Pages. 21–27, 1967.

30. Maron, M. E., "Automatic Indexing: An Experimental Inquiry," J. ACM, Vol. 8, Issue. 3, Pages. 404–417, 1961.

31. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., Classification And Regression Trees. Routledge, 2017.

32. Breiman, L., "Random forests," Mach. Learn., Vol. 45, Issue. 1, Pages. 5–32, 2001.

33. Breiman, L., "Bagging predictors," Mach. Learn., Vol. 24, Issue. 2, Pages. 123–140, Aug. 1996.

34. Wright, M. N., and Ziegler, A., "ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R," J. Stat. Softw., Vol. 77, Issue. 1, 2017.

35. Schapire, R. E., "The strength of weak learnability," Mach. Learn., Vol. 5, Issue. 2, Pages. 197–227, Jun. 1990.

36. Freund, Y., and Schapire, R. E., "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J. Comput. Syst. Sci., Vol. 55, Issue. 1, Pages. 119–139, Aug. 1997.

37. Friedman, J. H., "Greedy function approximation: A gradient boosting machine," Ann. Stat., Vol. 29, Issue. 5, Oct. 2001.

38. Vapnik, V. N., The Nature of Statistical Learning Theory. New York, NY: Springer New York, 1995.

39. Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning representations by back-propagating errors," Nature, Vol. 323, Issue. 6088, Pages. 533–536, Oct. 1986.

40. Chen, T., and Guestrin, C., "XGBoost," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, Vol. 13-17-Augu, Pages. 785–794.

41. Zou, H., and Hastie, T., "Regularization and variable selection via the elastic net," J. R. Stat. Soc. Ser. B Stat. Methodol., Vol. 67, Issue. 2, Pages. 301–320, 2005.