

Article

Not peer-reviewed version

Predicting Early-Stage Heart Failure Using Artificial Intelligence Techniques

Hamid Akram and [Noor ul Amin](#) *

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1137.v1

Keywords: heart failure; machine learning; prediction; classification; healthcare; supervised learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Predicting Early-Stage Heart Failure Using Artificial Intelligence Techniques

Hamid Akram ¹ and Noor ul Amin ^{2,*}

Department of Software Engineering, University of Sialkot, Sialkot, Pakistan

School of Computer Science, Taylor's University, Subang Jaya, Malaysia

* Correspondence: nooraminnawab@gmail.com

Abstract

Heart failure is an incurable condition in which the heart gradually loses its ability to pump blood effectively. It is a growing global health concern affecting millions of people worldwide. The risk of heart failure increases with age, highlighting the need for machine learning models capable of predicting heart failure at an early stage. Early predictions can help reduce disease progression, lower hospitalization rates, and improve patients' quality of life. The primary objective of this study is to predict patients in the early stages of heart failure using machine learning techniques based on health-related attributes. By leveraging the Cleveland dataset, which includes 13 key health features, our system predicts heart failure with high precision, enabling early intervention and more effective treatment planning. These models were tested and evaluated using standard performance metrics. Among them, the Random Forest classifier, implemented using RapidMiner, achieved the highest accuracy of 92.16%, outperforming other models in predictive capability.

Keywords: heart failure; machine learning; prediction; classification; healthcare; supervised learning

1. Introduction

Heart failure is a major health concern in which the heart struggles to pump blood effectively[1]. This condition can lead to various complications, frequent hospitalizations, and even death. In the United States alone, approximately 6.2 million people are affected by heart failure, with the condition being more prevalent among older adults[2,3]. Several risk factors contribute to the development of heart disease, including high blood pressure, existing heart conditions, and a history of heart attacks[4].

Early symptoms of heart failure may include shortness of breath, fatigue, and swelling in the legs and ankles [7]. If not detected at an early stage, the disease can progress and result in severe outcomes, such as heart attacks and reduced quality of life. Currently, doctors rely on clinical tests and checkups to diagnose heart disease, but these methods may not always be sufficient for early detection [9].

This highlights the need for more effective approaches to identify heart conditions before they become critical. Machine learning, a modern technological advancement, offers promising solutions by analyzing large datasets to detect patterns associated with disease. By using a patient's health information and lifestyle factors, machine learning can assist in the early prediction of heart disease. As the number of cases continues to rise, it becomes increasingly important to adopt improved methods for early diagnosis and prevention.

Figure 1 provides a visual comparison between a healthy heart and a diseased heart, illustrating how a heart attack can occur due to a blood clot blocking blood flow in a coronary artery affected by atherosclerosis.

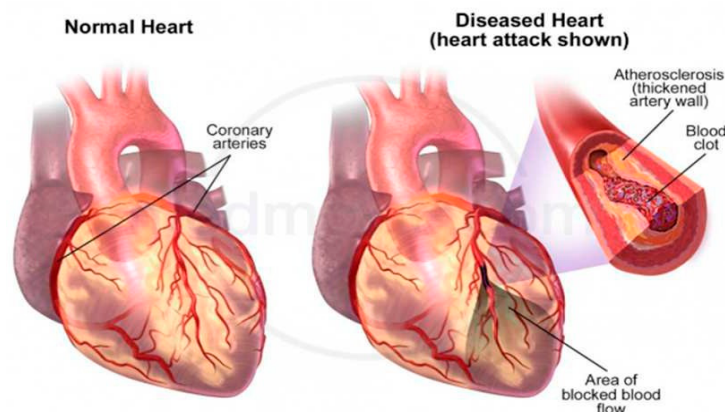


Figure 1. Normal Heart vs Disease Heart.

2. Literature Review

[1] address the challenges of heart disease prediction, particularly focusing on issues related to low accuracy and scalability. Their research compares different machine learning algorithms [11]. By applying LR, KNN and RF to the UCI dataset, which contains 13 medical attributes, the authors achieved an accuracy of 88.52% using the KNN model.

[2] investigate machine learning techniques for heart attack prediction [12]. This disease causes a significant number of deaths worldwide every year. In their paper, they study various machine learning approaches, including SVM, KNN, Decision Trees (DT), and RF. The Random Forest model achieved an accuracy of 91%, showing its potential in early disease detection and treatment.

[3] aim to solve the critical challenge of heart disease prediction, a major global cause of death. They highlight previous limitations in machine learning approaches [13]. The authors developed a hybrid model called HRFLM (Hybrid Random Forest with Linear Model). Using Neural Networks, SVMs, and Decision Trees, they achieved an accuracy of 88.7%, while the HRFLM model achieved 89.01% accuracy, demonstrating improved performance.

[4] discuss the difficulties in early heart failure prediction. Early detection is essential to reduce the mortality rate. The authors used the Cleveland dataset with 14 attributes and applied models such as Random Forest, SVM, Naïve Bayes, and Decision Trees. They achieved a maximum accuracy of 84%.

[5] address the challenge of heart disease prediction using various machine learning algorithms. In their study, they used the UCI dataset with 14 attributes and trained models including KNN, Decision Tree, SVM, and Linear Regression. Among these, KNN achieved the highest accuracy of 87%, followed by SVM (83%) and Decision Tree (79%). They concluded that KNN was the best-performing algorithm for their dataset.

[6] focus on predicting mortality in patients with heart failure. They found that traditional models like MAGGIC and ADHERE were inadequate in capturing the complete clinical picture. To address this, they developed a new machine learning-based model called MARKER-HF, using eight attributes. This model achieved an AUC of 0.81–0.84, outperforming traditional models like NT-proBNP. Their study emphasizes the benefit of machine learning for both patients and clinicians in early detection and personalized treatment.

[7] This paper describes how heart failure often follows a heart attack and is influenced by 13 contributing factors. The HRFLM model was used and achieved an accuracy of 88.7%, showing improved reliability and accuracy over older models. This research demonstrates the model's effectiveness in early detection of heart disease.

[8] This study discusses the challenges of diagnosing and predicting heart failure, particularly due to the complexity of patient data. The authors advocate for machine learning techniques over

traditional models such as the Seattle Heart Failure Model. Their results show that machine learning provides more accurate and reliable predictions of heart failure.

3. Proposed Methodology

Machine learning offers positive advancements in different fields, including healthcare. In this research, different machine learning algorithms are used to analyze datasets related to heart failure prediction. To solve the issue, it is important that the datasets used have no missing value and are well organized. The RapidMiner tool was chosen for applying machine learning techniques to the dataset. By using different types of algorithms, our goal is to achieve high accuracy for early-stage prediction of heart disease. Figure 2 shows our proposed framework.

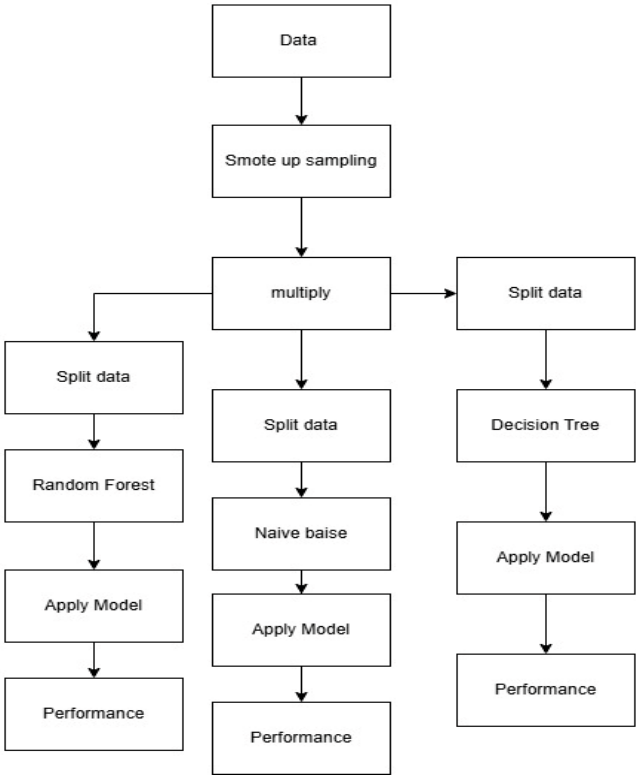


Figure 2. Framework.

3.1. Dataset

This dataset is downloaded from Kaggle, focusing on heart disease. It consists of 11 attributes that represent critical conditions for a patient. In this paper, it is studied how to predict heart failure early based on clinical data, biomarkers, genetics, and imaging results. The researchers identified that patients who start early treatment for heart disease have fewer chances of heart failure. This paper consists of different information to avoid risks and improve how we prevent and manage heart failure [8].

Prediction of cardiovascular disease is difficult because many risk factors are involved. To improve prediction over previous models, the author introduced a new method called Hybrid Random Forest with Linear Model (HRFLM). They used dataset attributes with 919 records. The dataset gives results on the basis of "Yes" or "No." It furnishes information based on attributes such as Age, RestingBP, Cholesterol, Resting ECG, etc.

3.2. Model Selection:

The selected model is based on higher accuracy compared to other models. There are a few steps to perform Random Forest on the dataset:

- A. Step 1: Import the dataset into RapidMiner.

- B. Step 2: Add SMOTE to balance the dataset.
- C. Step 3: Split the dataset into two parts: Training data and Testing data with a ratio of 80/20, as shown in Figure 3.

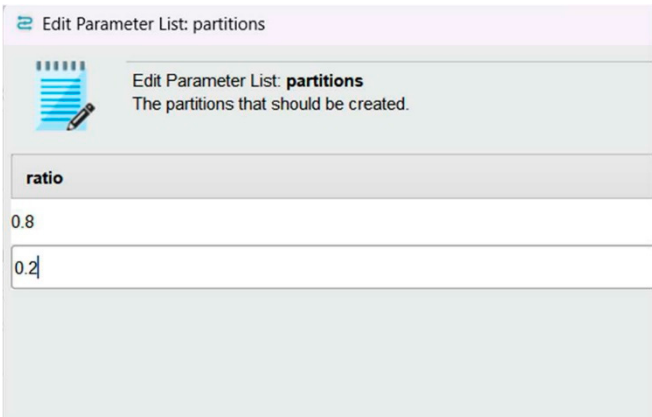


Figure 3. Splitting Data.

After splitting the data, several machine learning models were tested, such as Random Forest, KNN, Logistic Regression, Naive Bayes, and Decision Tree. By applying these classifiers, the Random Forest model gave higher accuracy than the others. In this paper, we choose Random Forest as the model to predict heart failure disease, as shown in Figure 4.

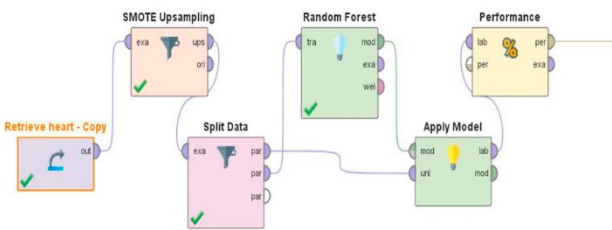


Figure 4. Overview of the model in Rapid Miner.

4. Results

In this section, we discuss the results of the classifiers used to predict heart disease, based on the findings of our experimental analysis. To optimize the accuracy of our model, we conducted experiments using various classifiers including Decision Tree, Logistic Regression, Naive Bayes, Random Forest, and K-Nearest Neighbors (KNN). We then compared the accuracy achieved by each model to determine the most effective approach. The results obtained from these algorithms are shown in Table 1.

Table 1. ACCURACY IN DIFFERENT CLASSIFIERS.

ALGORITHMS	ACCURACY (%)
DECISION TREE	85.47
NAIVE BAYES	84.26

RANDOM FOREST	92.16
LOGISTIC REGRESSION	83.13

The confusion matrix depicted in Figure 5 visually represents the performance of the Random Forest model in our system. It presents the class recall and class precision results. The confusion matrix serves as a crucial tool for evaluating a classification model’s performance by detailing outcomes such as true positives, true negatives, false positives, and false negatives. The results are compared with those of other authors using the same dataset, highlighting how our model achieves better accuracy. Table 2 shows a comparison of accuracy results from previous models that used different classifiers.

	true no	true yes	class precision
pred. no	91	5	94.79%
pred. yes	11	97	89.81%
class recall	89.22%	95.10%	

Figure 5. Splitting Data.

Table 2. Comparison of Results with Other Authors.

Comparison of results				
Author	Year	Data Set	Classification	Accuracy
[7]	2021	UCI	KNN	88.52%
[14]	2018	Cleveland	Random forest	91%
[15]	2019	UCI	HRFLM	89.01%
[16]	2020	Cleveland	SVC	84%

5. Conclusion and Future Work

This paper highlights the importance of early intervention to help patients at risk of heart failure. Using supervised machine learning techniques on the Cleveland dataset, we tested various classifiers. The highest accuracy of 92.16% was achieved by the Random Forest algorithm, outperforming the other methods used in RapidMiner. The dataset was first cleaned by removing irrelevant features and handling missing values, ensuring it was ready for model training. The models were then evaluated using performance metrics such as accuracy, recall, AUC, and the confusion matrix. This approach improves the efficiency and accuracy of heart failure prediction. With the help of such predictions, healthcare providers can act earlier, detect heart failure in its initial stages, and design better treatment plans.

References

1.

A. Pandey et al., “Biomarker-based risk prediction of incident heart failure in pre-diabetes and diabetes,” JACC: Heart Failure, vol. 9, no. 3, pp. 215–223, Mar. 2021, doi: 10.1016/j.jchf.2020.10.013.

2.

S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” IEEE Access, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

3. D. Jenča et al., "Heart failure after myocardial infarction: incidence and predictors," *ESC Heart Failure*, vol. 8, no. 1, pp. 222–237, Feb. 2021, doi: 10.1002/ehf2.13144.
4. E. D. Adler et al., "Improving risk prediction in heart failure using machine learning," *European Journal of Heart Failure*, vol. 22, no. 1, pp. 139–147, Jan. 2020, doi: 10.1002/ehf.1628.
5. S. K. Soni, ICE3-2020: International Conference on Electrical and Electronics Engineering, Feb. 14–15, 2020, IEEE, 2020.
6. Proceedings of the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), May 29–31, 2018, IEEE, 2018.
7. H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012072.
8. A. Mir et al., "A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques," *ESC Heart Failure*, Jul. 2024, doi: <https://doi.org/10.1002/ehf2.14942>.
9. W. Gouda, M. Almurafeh, M. Humayun, and N. Z. Jhanjhi, "Detection of COVID-19 based on chest X-rays using deep learning," *Healthcare (Switzerland)*, vol. 10, no. 2, art. no. 343, 2022, doi: 10.3390/healthcare10020343.
10. N. Zaman, T. J. Low, and T. Alghamdi, "Energy efficient routing protocol for wireless sensor network," in *Proc. Int. Conf. Advanced Communication Technology (ICACT)*, 2014, pp. 808–814, doi: 10.1109/ICACT.2014.6779072.
11. M. Lim, A. Abdullah, N. Z. Jhanjhi, M. Khurram Khan, and M. Supramaniam, "Link prediction in time-evolving criminal network with deep reinforcement learning technique," *IEEE Access*, vol. 7, pp. 184797–184807, 2019, doi: 10.1109/ACCESS.2019.2958873.
12. M. A. Hossain, S. K. Ray, and J. Lota, "SmartDR: A device-to-device communication for post-disaster recovery," *Journal of Network and Computer Applications*, vol. 171, art. no. 102813, 2020, doi: 10.1016/j.jnca.2020.102813.
13. H. Shahid, H. Ashraf, A. Ullah, S. S. Band, and S. Elnaffar, "Wormhole attack mitigation strategies and their impact on wireless sensor network performance: A literature survey," *International Journal of Communication Systems*, vol. 35, no. 16, art. no. e5311, 2022, doi: 10.1002/dac.5311.
14. Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology*, 7(2.8), 684-687.
15. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7, 81542-81554.
16. Almulhim, M., Islam, N., & Zaman, N. (2019). A lightweight and secure authentication scheme for IoT based e-health applications. *International Journal of Computer Science and Network Security*, 19(1), 107-120.
17. Zaman, N., Low, T. J., & Alghamdi, T. (2014, February). Energy efficient routing protocol for wireless sensor network. In *16th international conference on advanced communication technology* (pp. 808-814). IEEE.
18. Azeem, M., Ullah, A., Ashraf, H., Jhanjhi, N. Z., Humayun, M., Aljahdali, S., & Tabbakh, T. A. (2021). Fog-oriented secure and lightweight data aggregation in iomt. *IEEE Access*, 9, 111072-111082.
19. Ahmed, Q. W., Garg, S., Rai, A., Ramachandran, M., Jhanjhi, N. Z., Masud, M., & Baz, M. (2022). Ai-based resource allocation techniques in wireless sensor internet of things networks in energy efficiency with data optimization. *Electronics*, 11(13), 2071.
20. Khan, N. A., Jhanjhi, N. Z., Brohi, S. N., Almazroi, A. A., & Almazroi, A. A. (2022). A secure communication protocol for unmanned aerial vehicles. *CMC-Computers Materials & Continua*, 70(1), 601-618.
21. [21]Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
22. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2023). An intelligent healthcare system using IoT in wireless sensor network. *Sensors*, 23(11), 5055.

23. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications (pp. 49-64). IGI Global.
24. Sujatha R., Aarthy S.L., Chatterjee J.M., Alaboudi A., Jhanjhi N.Z. (2021). A Machine Learning Way to Classify Autism Spectrum Disorder. International Journal of Emerging Technologies in Learning, 16(6), 182-200. <https://doi.org/ijet.v16i06.19559>
25. Cheti I.A., Humayun M., Sama N.U., Jhanjhi N.Z. (2020). Evolution, Mitigation, and Prevention of Ransomware. 2020 2nd International Conference on Computer and Information Sciences, ICCIS 2020, , . <https://doi.org/ICCIS49240.2020.9257708>
26. Humayun M., Jhanjhi N.Z., Niazi M., Amsaad F., Masood I. (2022). Securing Drug Distribution Systems from Tampering Using Blockchain. Electronics (Switzerland), 11(8), . <https://doi.org/electronics11081195>
27. Zahra F., Jhanjhi N.Z., Brohi S.N., Khan N.A., Masud M., AlZain M.A. (2022). Rank and Wormhole Attack Detection Model for RPL-Based Internet of Things Using Machine Learning. Sensors, 22(18), . <https://doi.org/s22186765>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.