Optimal Tree Depth in Decision Tree Classifiers for Predicting Heart Failure Mortality

Tsehay Admassu Assegie[1*] , Ahmed Elaraby[2]

[1] Department of Computer Science, College of Engineering and Technology, Injibara University, 40 Injibara, Ethiopia

[2] Department of Computer Science, Faculty of Computers and Information, South Valley University, 83523 Qena, Egypt

* Corresponding Author Email: tsehayadmassu2006@gmail.com

Citation: Assegie, T. A. & Elaraby, A. (2023). Optimal tree depth in decision tree classifiers for predicting heart failure mortality. *Healthcraft Front.*, *1*(1), 58-66. https://doi.org/10.56578/hf010105.

Abstract: The depth of a decision tree (DT) affects the performance of a DT classifier in predicting mortality caused by heart failure (HF). A deeper tree learns complex patterns within the data, theoretically leading to better predictive performance. A very deep tree also leads to overfitting, because the model learns the training data rather than generalize to new and unseen data, resulting in a lower classification performance on test data. Similarly, a shallow tree does not learn much of the complexity within the data, leading to underfitting and a lower performance. The pruning method has been proposed to set a limit on the maximum tree depth or the minimum number of instances required to split a node to reduce the computational complexity. Pruning helps avoid overfitting. However, it does not help find the optimal depth of the tree. To build a better-performing DT classifier, it is crucial to find the optimal tree depth to achieve optimal performance. This study proposed cross-validation to find the optimal tree depth using validation data. In the proposed method, the cross-validated accuracy for training and test data is empirically tested using the HF dataset, which contains 299 observations with 11 features collected from the Kaggle machine learning (ML) data repository. The observed result reveals that tuning the DT depth is significantly important to balance the learning process of the DT because relevant patterns are captured and overfitting is avoided. Although cross-validation techniques prove to be effective in determining the optimal DT depth, this study does not compare different methods to determine the optimal depth, such as grid search, pruning algorithms, or information criteria. This is the limitation of this study.

Keywords: Machine learning (ML); Classification; Performance optimization; Heart failure (HF)

1. Introduction

Researchers have proposed various automated ML systems for predicting the risk of HF (Abdualgalil et al., 2022; Furizal et al., 2023). ML has provided intriguing new opportunities in improving patient outcomes in the medical healthcare field (Ali et al., 2023). Random forest, K-nearest neighbors (KNN), and support vector machine (SVM) classifiers have been proposed for predicting mortality caused by HF (Javeed et al., 2023). Experimental results demonstrate that SVM with linear kernel has an accuracy of 90.74% in predicting cardiac mortality, outperforming DT and KNN classifiers.

Furthermore, Shukur & Mijwi (2023) compared the performance of different ML techniques for HF diagnosis. The comparative analysis shows that the SVM model achieves 96% accuracy on the Cleveland clinical dataset. The result also suggested that the artificial neural network achieved 95% accuracy in predicting death occurrences due to cardiac failure.

Similarly, De Lio et al. (2023) and Mahmud et al. (2023) evaluated the performance of various ML techniques (e.g. DT, KNN, and light gradient boosting) for predicting cardiac failure death occurrences. The evaluation shows that random forest predicts HF more effectively than other ML models, with an accuracy of 87%.

Additionally, deep learning has also been used for predicting congestive HF (Rahman et al., 2023). It is usually difficult to identify HF at the early stage. In addition, HF brings other health complications at a later stage. Thus,

it is paramount to provide a more accurate and timely means to predict the severity of HF. ML enhances the prediction of HF (Kerexeta et al., 2023). Various optimization techniques, such as grid search and parameter tuning, have been highlighted as a method for enhancing the performance of ML for predicting HF.

DT is one of the most widely used algorithms for classification tasks (Goretti et al., 2022). For instance, Chandrasekhar & Peddakrishna (2023) developed an extra tree-based intelligent model for predicting HF. Predicting HF death occurrences improves patient outcomes as it helps identify at-risk HF patients. Similarly, Chen et al. (2023) predicted HF death occurrences using an extra tree classifier. The experimental result shows that this model is very helpful to clinicians in prioritizing patient care.

The performance evaluation of the DT, logistic regression, random forest, naïve Bayes, and SVM shows that the DT outperforms other ML models (Alotaibi, 2019). Furthermore, the review presented in the study shows that the DT model obtains 93.19% accuracy in HF survival prediction.

In a research conducted by Ghiasi et al. (2020), classification and regression trees (CART) were employed to automate the diagnosis of HF. The CART-based classifier achieved 98.61% accuracy for distinguishing between death occurrences and survival caused by HF. Although the overall accuracy was good, the classifier was not tested on the positive class for its accuracy.

Senan et al. (2021) proposed a correlation coefficient-based method to select optimal features for improving the performance of various ML algorithms (e.g. SVM, and KNN) for predicting HF. The performance of SVM and KNN can be improved to 95% with correlation-based feature selection. Although feature selection improved the performance of SVM and KNN, the impact of depth on the DT algorithm was not presented in their study.

Furthermore, in predicting HF events using an ML model, Tragante et al. (2022) compared the performance of different ML models (e.g. SVM, DT, and random forest). The performance comparison shows that the DT achieves 93.19% outperforming other models, such as random forest, logistic regression, and SVM. However, the effect of depth on the DT's performance was not highlighted.

Several studies (Mpanya et al., 2023; Pedro & Sá nchez, 2023; Sabovčik et al., 2022) and literature reviews that have been conducted to determine the optimal depth for a DT. In general, a optimal-depth DT can strike a balance between capturing important features and relationships in the data and avoiding overfitting (Dangare & Apte, 2012; Penny-Dimri et al., 2023). Therefore, it is an important task to find the optimal depth to build an accurate and reliable model (Beunza et al., 2019; Jang et al., 2023).

The DT depth is crucial because it directly influences the complexity and generalization ability of the model (Tong et al., 2023; Ayon et al., 2020). A deeper tree captures more complex relationships in the data, but it may also overfit the training data and perform poorly on new and unseen data. On the other hand, a shallower tree may not capture all the important features and relationships in the data, leading to underfitting and poor performance.

The DT classifier has inherent problems, with a lower value of the maximum tree depth leading to overfitting and a higher value leading to more computational time. This research highlights the importance of considering the DT depth in building effective classifiers for HF death prediction, because it is crucial for achieving a high accuracy.

This study provides valuable insights into the impact of tree depth on DT classifier performance and emphasizes the need for carefully considering this factor in predictive modeling for HF death events. Overall, this research aims to investigate the answers to three research questions: What is the optimal DT depth for the test set in HF prediction? What is the influence of depth on the performance of DTs? How can the efficiency of the DT be improved for HF prediction?

2. Method

The task of HF based on the ML model generally mainly includes four steps (Assegie et al., 2023; Austina et al., 2013; Awan et al., 2019; Pudjihartono et al., 2022). After data collection and pre-processing, training and testing data is split. Then after training the model on the training data, the final step involves validating the model on the test set. This study investigated the influence of the tree depth to predict the death event of HF patients using a dataset, which publicly available in the Kaggle data repository.

A DT model was trained on the collected dataset after splitting the dataset into training (70%) and testing (30%) dataset. The study employed various techniques, such as cross-validation, to test and evaluate different tree depths on validation data. It was important to carefully tune the tree depth to balance between capturing relevant patterns and avoiding overfitting. The flowchart in Figure 1 illustrates a general overview of the study.
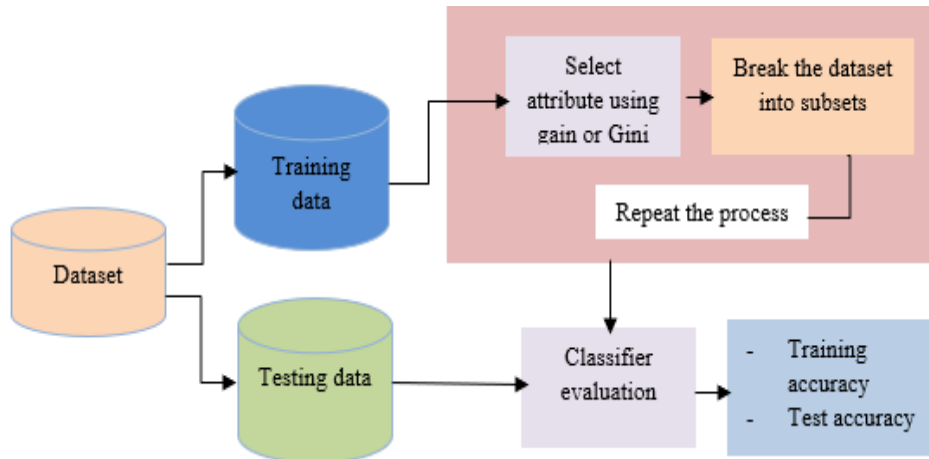
Figure 1. Flowchart for developing a DT classifier

2.1 Data Collection

2.1.1 Dataset description

The HF dataset contains 299 samples, of which 203 died, and 96 survived. The samples are described by 11 representative features, such as age, previous history of anemia, creatinine phosphokinase, ejection fraction, blood pressure, number of platelets, serum creatinine, serum sodium, sex, and smoking history of a patient. Table 1 describes the features used to describe each observation in the HF dataset. A statistical summary of the features is illustrated in Table 1.

Table 1. Explanation of the HF dataset

Number of Observations	Number of Classes	Number of Survivals	Number of Deaths	Number of Input Features
299	2	96	203	11

The dataset contains information on HF death events. The input features of the dataset include HF patient information, such as age, sex, serum creatinine, serum sodium, blood pressure, ejection fraction, creatinine phosphokinase, and platelets. Moreover, other previous historical information, such as anemia, and diabetes, are included in the input feature.

2.1.2 Statistical analysis of HF features

The statistical analysis includes the z-test and p-test for each independent feature in the HF dataset. These statistical tests are important for gaining insight into the significance of each HF input feature, thereby developing a better-performing DT model to predict the death event. The significance test using the z-test and p-test is illustrated in Table 2.

Table 2. Statistical description of independent features

Input Feature	Description	z-Test	p-Test
Age	Continuous	4.98	<0.005
Anaemia	Boolean	2.12	0.03
Creatinine phosphokinase	Continuous, mcg/L	2.23	0.03
Diabetes	Boolean	0.63	0.53
Ejection fraction	Continuous	-4.67	<0.005
High blood pressure	Boolean	2.20	0.03
Platelets	Continuous	-0.41	0.68
Serum creatinine	Serum creatinine mg/dL	4.58	<0.005
Serum sodium	Serum creatinine mg/dL	-1.90	0.06
Sex	Male or female	-0.94	0.35

The significance analysis for each of the input features is summarized in Table 2. It is observed from Table 2 that the age of the patient, ejection fraction, and serum creatinine have *p*-values <0.005. Based on the significance test, age, serum creatinine, and ejection fraction have much significance on the predictive performance of the DT.

2.1.3 Mean value analysis

The arithmetic mean is one of the most intuitive measures of central tendency (Javid et al., 2020). The variable of size n consists of the values (X1, X2, ... Xn). The arithmetic mean of this data is defined as the formula given in Eq. (1):

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{n} X_i \tag{1}$$

where, *N* is the number of total values of a given data point or sample.

The mean values of the HF features are revealed in Figure 2. As observed from Figure 2, the mean values of age, phosphokinase, and blood pressure of the HF patients are higher for death caused by HF. However, the ejection fraction and serum are lower. Thus, it can be concluded that HF patient with lower blood ejection and serum creatine has a higher chance of death.

	Death event			Survival
Age	65.22		Age	58.76
Anaemia	0.48		Anaemia	0.41
Phosphokinase	670.20		Phosphokinase	540.05
Diabetes	0.42		Diabetes	0.42
e_fraction	33.47		e_fraction	40.27
blood_pressure	0.41		blood_pressure	0.33
platelets	256381.04		platelets	266657.49
s_creatinene	1.84		s_creatinene	1.18
s_sodium	135.38		s_sodium	137.22
sex	0.65		sex	0.65
smoking	0.31		smoking	0.33
time	70.89		time	158.34
DEATH_EVENT	1.00		DEATH_EVENT	0.00
	Mean			Mean

Figure 2. Mean values of HF features for survival and death events

2.1.4 Correlation of each feature with the death event

The correlation value of the HF feature varies from –1 to 1. When it is closer to 1, it means that a strong positive relationship exists. For example, the median age value tends to increase the probability of a death event. When the coefficient is closer to –1, it means that there is a strong negative relationship (Lu et al., 2022). Figure 3 shows correlation coefficients, highlighting the relationships between various HF features and the likelihood of death events.

The correlation between the dependent variable (death event), and the independent variables or the HF input feature is given by the formula in Eq. (2).

$$r = \frac{\sum (X_i - \overline{X_i})(Y_i - \overline{Y_i})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X_i})^2 + (Y_i - \overline{Y_i})^2}} \tag{2}$$

where, *r* signifies the Pearson correlation coefficient, *N* signifies the number of data points in the HF dataset, X_i signifies values of the variable *X* in the HF dataset, Y_i signifies values of the variable *Y* in the HF dataset, $\overline{X}$ denotes the mean of variable *X*, and $\overline{Y}$ signifies the mean of the values of variable in heart disease dataset.

As indicated in Figure 3, serum creatinine, age, and blood pressure are strongly correlated to death events. In contrast, time, ejection fraction, serum sodium, platelets, smoking, and sex have a negative correlation to death event caused by HF.
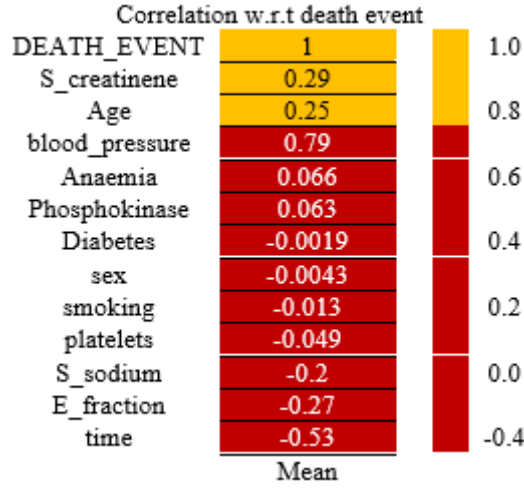
Figure 3. Correlation of HF with death events

2.1.5 Evaluation criteria

Accuracy is the most important evaluation metric for the effectiveness of ML. It helps assess the effectiveness of DT (Shehzadi et al., 2022). To test the improbability of the DT, prediction probability is employed as an efficiency. Another useful metric used for the assessment of the model is the accuracy score and receiver operating characteristics (ROC) curve. A classifier's accuracy and ROC curve are defined as follows.

$$Accuracy = \frac{T}{N} \tag{3}$$

where, T is the correctly predicted test examples, and N is the total samples considered in testing. In the HF dataset, there are more types of errors because one class occurs more frequently than the other.

This is very common in practice, especially in medical datasets, where the data samples in the positive class are lower than the negative (Suresh et al., 2022). In such cases, accuracy might not sufficiently describe the performance score of a classification model. Accuracy does not sufficiently quantify the efficiency of the predictive model in imbalanced classification tasks (Qian et al., 2022). Therefore, the ROC curve is employed as a performance metric for evaluating the classification model on HF death prediction. In addition, a confusion matrix is also employed as an alternative performance measure to provide better results in evaluating and selecting a predictive model for the imbalanced classification task. The matrix is also commonly used to evaluate the effectiveness of the ML model for HF prediction (Alizadehsani et al., 2019). The matrix is a two-dimensional array, where the rows correspond to the true class (ground truth) and the column corresponds to the predicted class.

The observations correctly predicted that HF death event is true positive (TP) and the observations correctly predicted as non-HF death is true negative (TN). In contrast, the observations incorrectly classified as HF patient death are false positive (FP) and those incorrectly predicted as not survived HF patient are true negative (TN).

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

The true positive rate (TPR) indicates the portion of HF death events among correctly predicted ones. In dissimilarity, the false positive rate (FPR) indicates the number of HF death events among the HF survival predictions. The FPR is given as follows:

$$FPR = \frac{FP}{TP + TN} \tag{5}$$

The precision defines the fraction of TP, among all instances that the DT has categorized as positive: NTP and number of false positives (NFP). The value is determined with the following formula.

$$\text{Precsion} = \frac{\text{NTP}}{\text{NTP} + \text{NFP}} \tag{6}$$

3. Results and Discussion

In this section, the proposed model is evaluated on the test set using training, test, and cross-validation accuracy on the HF dataset. In summary, the impact of DT depth on accuracy is an important consideration in ML. Finding the right balance between underfitting and overfitting is crucial for building effective classification models. Thus, different tree depths need to be carefully evaluated using cross-validation techniques to maximize accuracy. This study emphasizes the trade-off between overfitting and underfitting when determining the optimal DT depth. Techniques, such as pruning and cross-validation, are suggested to find the right balance that achieves the best predictive power.

3.1 Effect of Depth on DT Classifier

Additionally, the impact of tree depth on interpretability is highlighted, emphasizing the need to consider both predictive power and interpretability when determining the optimal DT depth. Overall, the reviews stress the importance of finding the optimal depth to maximize predictive power while maintaining interpretability.

As shown in Figure 4, the maximum DT depth is an important attribute contributing to the performance of the DT, significantly affecting the training, testing, and cross-validation accuracy. Table 3 delineates the relationship between the depth of the tree and its corresponding performance metrics.

Table 3. Performance of DT vs depth of the tree

Depth	Training Accuracy	Testing Accuracy	Cross-Validation Accuracy
1	75.12	72.22	73.72
2	87.08	80	81.36
3	91.87	83.33	86.36
4	94.26	85.56	86.13
5	95.69	81.11	85.18
6	97.61	82.22	87.08
7	100	84.44	84.23
8	100	82.22	85.18
9	100	81.11	85.66
10	100	83.33	85.66
11	100	83.33	85.66
12	100	83.33	85.66
13	100	83.33	85.66
14	100	82.22	85.66
15	100	82.22	85.66
16	100	82.22	85.66
17	100	82.22	85.66
18	100	82.22	85.66
19	100	82.22	85.66
20	100	82.22	85.66

In this case, the result illustrated in Figure 4 suggests that a maximum depth value between 3 and 7 is appropriate, confirming higher values of cross-validation accuracy while reducing the overfitting probability. A depth value of more than 7 results in overfitting with training accuracy closer to 100%, and lower test and cross-validation accuracy in HF death event prediction. The increase in the DT depth leads to better performance on the training data, as the model captures more intricate patterns and relationships within the data. However, deeper trees are also more prone to overfitting, because they learn to memorize the training data rather than generalize to new and unseen data, leading to poor performance on test or validation data. The lower-depth trees may not capture all the nuances of the data and underfit, leading to lower accuracy and predictive power. Therefore, it is crucial to find the optimal DT depth to achieve the best performance. It is evident from Figure 4 that the optimal depth value can be selected through cross-validation.
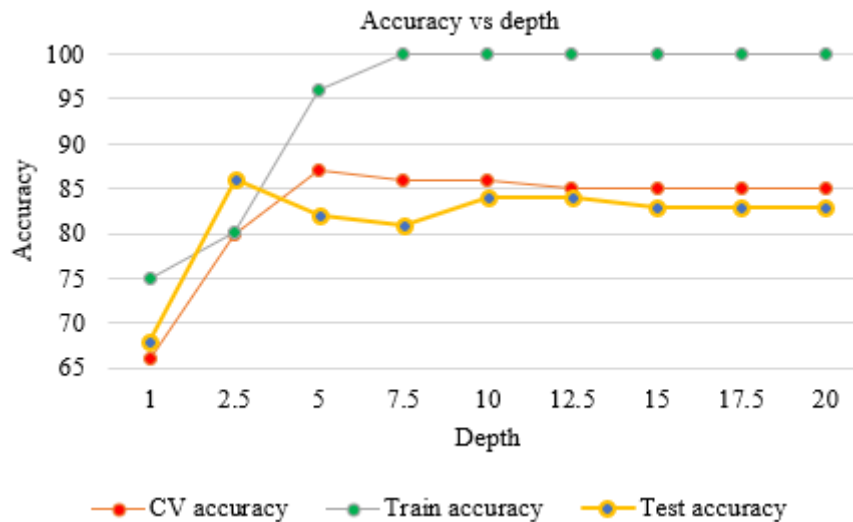
Figure 4. Accuracy vs depth

4. Conclusions

In conclusion, it is crucial to find the optimal DT depth to achieve high accuracy and generalizability in predicting HF death events. This research highlights the importance of incorporating DT depth into building effective classifiers for HF death prediction, with depth range of 1 to 20 being considered. A deeper tree captured more intricate patterns and interactions within the data, leading to better predictive performance. However, there was also an overfitting risk, because the model memorized the training data and did not generalize well to new data. It was found that the DT depth significantly affected the predictive performance. Overall, this study provides valuable insights into the impact of DT depth on accuracy and emphasizes the need for careful consideration of this parameter in building effective classification models. It is also recommended that advanced pruning techniques and ensemble methods can be further studied to improve the accuracy of DTs across different datasets and applications. However, apart from the dataset used in this study, other datasets need to be utilized. Furthermore, it is recommended that other pre-processing techniques and larger depth space than the one considered in this study can be used in future studies, thereby confirming the findings of this study.

Ethical Approval

This study adheres to strict ethical guidelines, ensuring the rights and privacy of participants. Informed consent was obtained from all participants, and personal information was protected throughout the study. The methodology and procedures of this research have been approved by the appropriate ethics committee. Participants were informed of their rights, including the right to withdraw from the study at any time. All collected data is used solely for this research and is stored and processed securely and confidentially.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abdualgalil, B., Abraham, S., & Ismael, W. M. (2022). Early diagnosis for dengue disease prediction using efficient machine learning techniques based on clinical data. *J. Robotics Control, 3*(3), 257-268. https://doi.org/10.18196/jrc.v3i3.14387.

Ali, M. M., Al-Doori, V. S., Mirzah, N., Hemu, A. A., Mahmud, I., Azam, S., Al-tabatabaie, K. F., Ahmed, K., Bui, F. M., & Moni, M. A. (2023). A machine learning approach for risk factors analysis and survival prediction of heart failure patients. *Healthc. Anal., 3*, 100182. https://doi.org/10.1016/j.health.2023.100182.

Alizadehsani, R., Roshanzamir, M., Abdar, M., Beykikhoshk, A., Khosravi, A., Panahiazar, M., Koohestani, A., Khozeimeh, F., Nahavandi, S., Sarrafzadegan, N., & Acharya, U. R. (2019). A database for using machine learning and data mining techniques for coronary artery disease diagnosis. *Sci. Data, 6*(1), 227. https://doi.org/10.1038/s41597-019-0206-3.

Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *Int. J. Adv. Comput. Sci. Appl., 10*(6), 262–268. https://doi.org/10.14569/IJACSA.2019.0100637.

Assegie, T. A., Elanangai, V., Paulraj, J. S., Velmurugan, M., & Devesan, D. F. (2023). Evaluation of feature scaling for improving the performance of supervised learning methods. *Bull. Electr. Eng. Inform., 12*(3), 1833–1838. https://doi.org/10.11591/eei.v12i3.5170.

Austin, P. C., Tu, J. V., Ho, J. E., Levy, D., & Lee, D. S. (2013). Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. *J. Clin. Epidemiol., 66*(4), 398-407. http://doi.org/10.1016/j.jclinepi.2012.11.008.

Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., & Dwivedi, G. (2019). Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. *ESC Heart Fail., 6*(2), 428-435. https://doi.org/10.1002/ehf2.12419.

Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020). Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. *IETE J. Res.*, 68(4), 2488-2507. https://doi.org/10.1080/03772063.2020.1713916.

Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., Hurtado, C., & Landecho, M. F. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). *J. Biomed. Inform., 97*, 103257. https://doi.org/10.1016/j.jbi.2019.103257.

Chandrasekhar, N. & Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes, 11*(4), 1210. https://doi.org/10.3390/pr11041210

Chen, S. Y., Hu, W. W., Yang, Y. H., Cai, J. X., Luo, Y. Q., Gong, L. M., Li, Y. M., Si, A. M., Zhang, Y. X., Liu, S. T., Mi, B. B., Pei, L. L., Zhao, Y. L., Chen, F. Y. (2023). Predicting six-month re-admission risk in heart failure patients using multiple machine learning methods: A study based on the Chinese heart failure population database. *J. Clin. Med., 12*(3), 870. https://doi.org/10.3390/jcm12030870.

Dangare, C. S. & Apte, S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. *Int. J. Comput. Appl., 47*(10), 44-48.

De Lio, F., Andreis, A., De Lio, G., Bellettini, M., Pidello, S., Raineri, C., Gallone, G., Alunni, G., Frea, S., Imazio, M., Castagno, D., & De Ferrari, G. M. (2023). Cardiac imaging for the prediction of sudden cardiac arrest in patients with heart failure. *Heliyon, 9*(7), e17710. https://doi.org/10.1016/j.heliyon.2023.e17710.

Furizal, A., Ma'arif, A., & Rifaldi, D. (2023). Application of machine learning in healthcare and medicine: A review. *J. Robotics Control, 4*(5), 621-631.

Ghiasi, M. M., Zendehboudi, S., & Mohsenipour, A. A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. Comput. *Methods Programs Biomed.*, *192*, 105400. https://doi.org/10.1016/j.cmpb.2020.105400.

Goretti, F., Oronti, B., Milli, M., & Iadanza, E. (2022). Deep learning for predicting congestive heart failure. *Electron., 11*(23), 3996. https://doi.org/10.3390/electronics11233996.

Jang, S. Y., Park, J. J., Adler, E., Eshraghian, E., Ahmad, F. S., Campagnari, C., Yagil, A., Greenberg, B. (2023). Mortality prediction in patients with or without heart failure using a machine learning model. *JACC: Adv., 2*(7), 100554. https://doi.org/10.1016/j.jacadv.2023.100554.

Javeed, A., Saleem, M. A., Dallora, A. L., Ali, L., Berglund, J. S., & Anderberg, P. (2023). Decision support system for predicting mortality in cardiac patients based on machine learning. *Appl. Sci., 13*(8), 5188. https://doi.org/10.3390/app13085188.

Javid, I., Alsaedi, A. K., & Ghazali, R. (2020). Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method. *Int. J. Adv. Comput. Sci. Appl., 11*(3), 540-551.

Kerexeta, J., Larburu, N., Escolar, V., Lozano-Bahamonde, A., Macía, I., Beristain Iraola, A., & Graña, M. (2023). Prediction and analysis of heart failure decompensation events based on telemonitored data and artificial intelligence methods. *J. Cardiovasc. Dev. Dis., 10*(2), 48. https://doi.org/10.3390/jcdd10020048.

Lu, H., Yao, Y., Wang, L., Yan, J., Tu, S., Xie, Y., & He, W. (2022). Research progress of machine learning and deep learning in intelligent diagnosis of the coronary atherosclerotic heart disease. *Comput. Math. Methods Med., 2022*, 3016532. https://doi.org/10.1155/2022/3016532.

Mahmud, I., Kabir, M. M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2023). Cardiac failure forecasting based on clinical data using a lightweight machine learning metamodeling. *Diagnostics, 13*(15), 2540. https://doi.org/10.3390/diagnostics13152540.

Mpanya, D., Celik, T., Klug, E., & Ntsinjana, H. (2023). Predicting in-hospital all-cause mortality in heart failure using machine learning. *Front. Cardiovasc. Med., 9*, 1032524. https://doi.org/10.3389/fcvm.2022.1032524.

Pedro, A. & Sánchez, M. (2023). Improvement of a prediction model for heart failure survival through explainable artificial intelligence. *Front. Cardiovasc. Med., 10*, 1219586. https://doi.org/10.3389/fcvm.2023.1219586.

Penny-Dimri, J. C., Bergmeir, C., Reid, C. M., Williams-Spence, J., Perry, L. A., & Smith, J. A. (2023). Tree-based survival analysis improves mortality prediction in cardiac surgery. *Front. Cardiovasc. Med., 10*, 1211600. https://doi.org/10.3389/fcvm.2023.1211600

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinform., 2*, 927312. https://doi.org/10.3389/fbinf.2022.927312.

Qian, X., Li, Y., Zhang, X., Guo, H., He, J., Wang, X., Yan, Y., Ma, J., & Guo, S. (2022). A cardiovascular disease prediction model based on routine physical examination indicators using machine learning methods: A cohort study. *Front. Cardiovasc. Med., 9*, 854287. https://doi.org/10.3389/fcvm.2022.854287.

Rahman, M. S., Rahman, H. R., Prithula, J., Chowdhury, M. E., Ahmed, M. U., Kumar, J., Murugappan, M., & Khan, M. S. (2023). Heart failure emergency readmission prediction using stacking machine learning model. *Diagnostics, 13*(11), 1948. https://doi.org/10.3390/diagnostics13111948

Sabovčik, F., Ntalianis, E., Cauwenberghs, N., & Kuznetsova, T. (2022). Improving predictive performance in incident heart failure using machine learning and multi-centre data. *Front. Cardiovasc. Med., 9,* 1011071. https://doi.org/10.3389/fcvm.2022.1011071.

Senan, E. M., Abunadi, I., Jadhav, M. E., & Fati, S. M. (2021). Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms. *Comput. Math. Methods Med., 2021*, 1-16. https://doi.org/10.1155/2021/8500314.

Shehzadi, S., Hassan, M. A., Rizwan, M., Kryvinska, N., & Vincent, K. (2022). Diagnosis of chronic ischemic heart disease using machine learning techniques. *Comput. Intell. Neurosci., 2022*, 1-9. https://doi.org/10.1155/2022/3823350.

Shukur, B. S. & Mijwi, M. M. (2023). Involving machine learning techniques in heart disease diagnosis: A performance analysis. *Int. J. Electr. Comput. Eng., 13*(2), 2177-2185. https://doi.org/10.11591/ijece.v13i2.pp2177-2185.

Suresh, T., Assegie, T. A., Rajkumar, S., & Kuma, N. K. (2022). A hybrid approach to medical decision-making: Diagnosis of heart disease with machine-learning model. *Int. J. Electr. Comput. Eng., 12*(2), 1831-1838. https://doi.org/10.11591/ijece.v12i2.pp1831-1838.

Tong, R., Zhu, Z., & Ling, J. (2023). Comparison of linear and non-linear machine learning models for time-dependent readmission or mortality prediction among hospitalized heart failure patients. *Heliyon, 9*(5), e16068. https://doi.org/10.1016/j.heliyon.2023.e16068.

Tragante, V., Pilbrow, A., & Poppe, K. (2022). Improving early detection and risk prediction in heart failure. *Front. Cardiovasc. Med., 9*, 930935. https://doi.org/10.3389/fcvm.2022.930935.