International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Check for updates

Decision trees' viability and efficiency in forecasting coronary heart disease

M Jawahar * and S Monika

Department of CSE, Malla Reddy University, Hyderabad, India.

International Journal of Science and Research Archive, 2024, 13(02), 3574-3579

Publication history: Received on 16 November 2024; revised on 26 December 2024; accepted on 28 December 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.2.2590

Abstract

Cardiovascular disease is one of the most prevalent disease categories and a major global health concern. For therapy to be effective and survival chances to increase, early identification is essential. A decision tree in this instance is a trustworthy Classification technique for calculating cardiovascular disease hazard as well as gathering data for the clinical decision-making process. However, it may be difficult to find a more accurate simulation of cardiovascular illness due to scalability issues. Therefore, the performance and scalability of decision trees for the prediction of cardiovascular illness are investigated in this study. The study assessed a decision tree's ability to predict cardiovascular disease. The model complexity, cross-validation score, ratings of training result from augmented training samples, and error matrix were castoff to measure the enactment. The proceeding claims that model of decision tree had an accuracy with 89.8% in predicting the occurrence of cardiovascular disease. As a result, decision trees are useful for predicting and identifying cardiovascular disease incidents among patients early on.

Keywords: Decision tree; Classification; Machine learning; Regression

1. Introduction

According to a recent study [1], machine learning techniques are now considered to be important in clinical decision-making. Preliminary clinical support from machine learning (ML) systems helps cardiovascular patients save time and money. Additionally, while diagnosing cardiovascular illness, machine-learning technologies help doctors make better decisions. Furthermore, a previous revision [2] confirmed that machine learning can identify cardiovascular disease with 96% accuracy. A comparison study of various supervised learning models and identified that SVM has 97% of accuracy rate in predicting cardiovascular sickness. When a support vector machine has higher accuracy than k-nearest neighbor, which is 91%.

1.1. Related work

In line with this, Mienye et al. [3] added weights to classification and regression trees (CART) to further enhance their performance. The study found that, with an accuracy of 93%, the proposed CART model performed better than additional CART oriented heart condition classification prototypes from the literature which is available. Though the anticipated model performed better in the sense of accuracy than the current models, only precision, accuracy as well as f-score remained compared. Furthermore, the modification does not include alternative collective prototypes. Additional considerations are important when assessing ML samples for the forecast of cardiovascular illness, such as error matrix, scalability, and complexity of model. [4].

Furthermore, a number of studies demonstrated that a collective model may forecast cardiovascular illness [5, 6]. The study also highlighted how important it is now to forecast cardiovascular illness in order to save lives. With an accuracy

* Corresponding author: M Jawahar

rate of 87.23%, the investigational outcomes for cardiovascular illness forecast by the multilayer perceptron showed promise in aiding in the early identification of cardiovascular illness.

M.Jan, A. A. Awan et al. [7] also proposed a collective prototypical by merging the prediction capacity of multiple prototypes to advance the machine learning model accuracy in-order to identifying cardiovascular illness. The projected collective learning method working naïve Bayesian, SVM, random decision forest, regression analysis, and neural network as base classifiers. This study used cardiovascular datasets from the UCI repository that were gathered from Cleveland and Hungarian. In terms of predicted accuracy, the study proved that an ensemble model is a better method.

Similarly, another collective model approach to cardiovascular disease risk prediction was proposed by Zaini and Awang [8]. According to the study, hybrid feature selection enhances the ensemble-stacking model's performance. Chi-square test and analysis of variance to cardiovascular illness characteristic advance the collective prototype for assessing the hazard of cardiovascular disease in a period of 10 years. These techniques yield a precision of 93.44% for a collective model approach according to a logistic regression prototype when the best features are chosen.

Extrapolative modelling for cardiovascular illness has become more significant in medical research as well as patient treatment [9]. Here precise calculation of cardiovascular hazard confirms that fitness repercussions also the hazard of cardiovascular illness are immediately taken into account along with machine learning oriented fitness hazard assessment prototypes. A lot of data is used in the contemporary healthcare sector. Laboratory tests, clinical reports, and patient descriptions are just a few of the many forms of patient data that are collected every day. Improvements in data storing technology also the rising usage of digital fitness record schemes in hospitals have made this likely to process and make judgments on larger amounts of data.

A study [10] evaluated cardiovascular risk by a variety of methods in ML, like decision tree, K-nearest neighbor, SVM, By an correctness score around 95%, SVM method attained superior to other models on Cleveland dataset. Similarly, a predictive paradigm for machine learning oriented cardiovascular illness prediction was proposed by Molla et al. [11]. The work trained decision trees, support vector machines, gradient boosting (XG), and XGB using the Cleveland cardiovascular disease datasets. According to the analysis, the decision tree model's accuracy increases when univariate feature collection is used. We identify 10 characteristics associated with cardiovascular disease and increase the decision tree's accuracy to 97.75 percent by using the univariate strategy for predictor selection.

Additionally, the artificial neural network's capacity to predict cardiac disease was enhanced by Sarra, A.M.Dinar et al. [12]. The artificial neural network oriented model to predict cardiovascular disease had a 93.44% accuracy rate. Artificial neural networks have an accuracy of 7.5% when related to support vector machine algorithm. It was shown that the generated model could be trained in less than a minute. A computer model for a patient's angiographic disease status has been developed combining deep learning and machine learning. Almulihi et al. evaluated a collaborative approach artificial intelligence model for the prediction of early heart illness [13]. Here this work advances the efficiency of the deep learning prototype through building a collaborative approach by deep learning and its basis. Here accuracy of the model mentioned as 98.42%.

A solid common voting collective model was created by Kumar and Vigneswari [14] to evaluate early heart disease patients. According to the study, grid search increases the suggested model's cross-validation accuracy. Furthermore, the majority voting ensemble model's accuracy has increased thanks to the preprocessing technique with scaling. All things considered, the built model generated predictions with an accuracy of 90% when pre-processing with grid search and scaling. Feature selection improves the ensemble-learning model's performance. M.A.Alim, S. Habib et al. [15], proposed permutation aspect significance and other tree-based feature selection improve the model of random forest algorithm accuracy in 1.5%. Additionally, this model predicts the incidence of cardiovascular illness more correctly when early evaluation techniques like normalization and misplaced value elimination are used. Monika S et al. [16] offer a remedy by creating a novel Internet of Things-based Maternal and Fetal Health Monitoring System which uses modern technology for improving prenatal care which examined by a diversity of ML models.

This examination was prompted via superior efficiency of ML methods within first learning published in publications [1] to [5]. Furthermore, the learning intends to measure the model's enactment because decision trees perform better in predicting cardiovascular illness. The study's objectives are to: a) examine the literature on machine learning methods for diagnosing cardiovascular disease, b) use various model evaluation results, and c) examine and assess the various decision trees' prediction abilities, including comparative analysis, model complexity, and confusion matrix, in adding to the correctness measure which is commonly used for prototype assessment.

2. The method

A discussion of the research timeline is given below. The first step is to collect data. The next step includes experimental data investigation using feature scaling and expressive statistics such as correlations and misplaced value assessment. The next phase debates developing a prototype using ensemble learning approaches. The next stage evaluates the model's efficiency using certain indicators like the complexity of model also error matrix. Here concluding step offerings the collection of a high execution prototype from the candidate model pool and the reference of an enhanced collective model for the primary analysis of cardiovascular disease.

A Cleveland UCI cardiovascular illness data source provided the dataset used in this investigation. One benchmark dataset is the UCI data repository. Numerous studies have previously confirmed its widespread application in machine learning research [17]–[19]. During the study's exploratory data analysis, descriptive statistics including standard deviation, mean, count, and maximum were used to examine the gathered data. Rotation estimation was castoff in this learning to measure how well collective learning prototype forecasted heart disease. Here process of separating dataset into a subclass of K-folds regards testing also the training is known as rotation estimation [20] to [22] is a method is commonly used for forecast models [23 - 26] to conclude in what way the built prototype handles unforeseen occurrences in training set. Following data collection, the feature is scaled and the data is examined for missing values. Thirdly, a decision tree method was handled to generate the classifier prototype. The last phase then entailed assessing the decision tree algorithm's efficiency by a number of processes, including rotation estimation precision and error matrix.

3. Result and discussion

Results in forecasting heart disease using collective learning models are presented in this section. Confusion matrix, Training time complexity, forecast precision are castoff to measure the efficiency in the collective model. In each model, the outcomes of standard performance measures are also displayed in this section, including the validation score and the model's scalability across a wider training sample.

Table.1 displays the decision tree model's measurement which is used to determine whether cardiac disease is present or not. As seen in Table.1, this model imperfectly forecast 41 out of the entire data handled in the process of testing. Furthermore, Table.1 determines that the model has correctness of 89% in training around 73% in forecasting real optimistic session as validation.

Table 1 Decision tree model's measurement

Measure	Value
Sensitivity	0.77
Specificity	0.87
Precision	0.89
Negative Predictive Value	0.73
False Positive Rate	0.13
False Discovery Rate	0.12
False Negative Rate	0.24
Accuracy	0.80
F1 Score	0.82
Matthews Correlation Coefficient	0.62

Figure.2 shows the fitting time for a varying number of training samples along with the scalability, cross-validation metrics. When related towards cross validation value the decision tree's training score produces better outcome. Additionally, up to 176 training samples, the cross-validation accuracy is excellent, and a training score is higher with 101 trials. Given that it matches the information in 0.18 seconds less than with a sample for training size of 176, the decision tree is a scalable model.

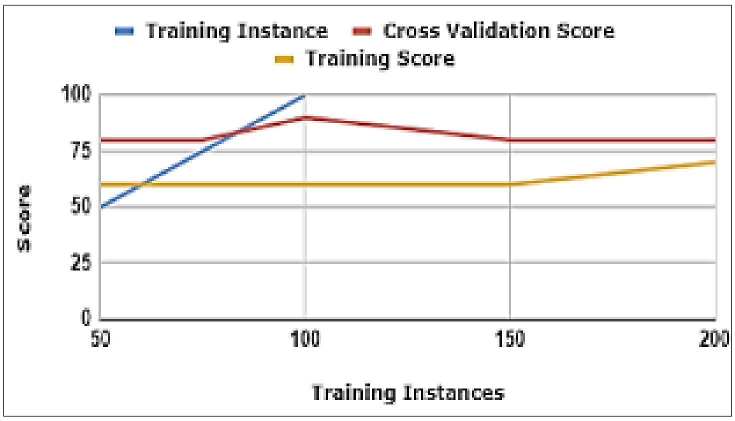

Figure 1 Decision tree model's scalability, training, and validation scores

4. Conclusion

This study examined the effectiveness of boosters in terms of time complexity by weighing scalability against ensemble learner performance. The study compared the decision tree metrics for the dataset on cardiovascular disease. As the trial progresses, it is discovered that the decision tree performs 83.3% cross validation. Researchers suggest more research on other datasets, including hypothyroidism and breast cancer, using other models, comprising SVM, collective learning approaches. Additionally, investigating feature scaling, feature selection strategies to address conflicts between complexity in the model and CV scores remains an expensive research topic. However, the study has some shortcomings here as well also it has numerous boundaries which are essential to be considered. Initially, this examination was restricted to a single dataset, which makes it impossible to generalize and compare it to other instances of cardiovascular disease. Second, only decision tree techniques were used for analysis; findings from other techniques can differ.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] P. Guleria, P. Naga Srinivasu, S. Ahmed, N. Almusallam, and F. K. Alarfaj, "XAI framework for cardiovascular disease prediction using classification techniques," Electronics, vol. 11, no. 24, 2022, doi: 10.3390/electronics11244086.

[2] B. S. Shukur and M. M. Mijwil, "Involving machine learning techniques in heart disease diagnosis: a performance analysis," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 2, pp. 2177-2185, Apr. 2023, doi: 10.11591/ijece.v13i2.pp2177-2185.

[3] I. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," Informatics in Medicine Unlocked, vol. 20, 2020, doi: 10.1016/j.imu.2020.100402.

[4] Y. Jiang et al., "Cardiovascular disease prediction by machine learning algorithms based on cytokines in kazakhs of china," Clinical Epidemiology, vol. 13, pp. 417–428, 2021, doi: 10.2147/CLEP.S313343.

[5] A. Alfaidi, R. Aljuhani, B. Alshehri, H. Alwadei, and S. Sabbeh, "Machine learning: assisted cardiovascular diseases diagnosis," International Journal of Advanced Computer Science and Applications, vol. 13, no. 2, pp. 135–141, 2022, doi: 10.14569/IJACSA.2022.0130216.

[6] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.

[7] M. Jan, A. A. Awan, M. S. Khalid, and S. Nisar, “Ensemble approach for developing a smart heart disease prediction system using classification algorithms,” Research Reports in Clinical Cardiology, vol. 9, pp. 33–45, 2018, doi: 10.2147/rrcc.s172035.

[8] N. A. M. Zaini and M. K. Awang, “Hybrid feature selection algorithm and ensemble stacking for heart disease prediction,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 2, pp. 158–165, 2023, doi: 10.14569/IJACSA.2023.0140220.

[9] T. A. Assegie, A. O. Salau, C. O. Omeje, and S. L. Braide, “Multivariate sample similarity measure for feature selection with a resemblance model,” International Journal of Electrical and Computer Engineering, vol. 13, no. 3, pp. 3359–3366, 2023, doi: 10.11591/ijece.v13i3.pp3359-3366.

[10] 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.

[11] S. Molla et al., “A predictive analysis framework of heart disease using machine learning approaches,” Bulletin of Electrical Engineering and Informatics, vol. 11, no. 5, pp. 2705–2716, 2022, doi: 10.11591/eei.v11i5.3942.

[12] R. R. Sarra, A. M. Dinar, and M. A. Mohammed, “Enhanced accuracy for heart disease prediction using artificial neural network,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 29, no. 1, pp. 375–383, 2023, doi: 10.11591/ijeecs.v29.i1.pp375-383.

[13] A. Almulihi et al., “Ensemble learning based on hybrid deep learning model for heart disease early prediction,” Diagnostics, vol. 12, no. 12, Dec. 2022, doi: 10.3390/diagnostics12123215.

[14] N. K. Kumar and D. Vigneswari, “A drug recommendation system for multi-disease in health care using machine learning,” Lecture Notes in Electrical Engineering, vol. 668, pp. 1–12, 2021, doi: 10.1007/978-981-15-5341-7_1.

[15] M. A. Alim, S. Habib, Y. Farooq, and A. Rafay, “Robust heart disease prediction: a novel approach based on significant feature and ensemble learning model,” 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies: Idea to Innovation for Building the Knowledge Economy, iCoMET 2020, 2020, doi: 10.1109/iCoMET48670.2020.9074135.

[16] Monika. S, H. Battana, M. Sangeetha, M. Shaik and M. Jawahar, "Analysis of Maternity and Child Health Care System Integrated with IoT and ML," 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2024, pp. 1792-1797, doi: 10.1109/ICACCS60874.2024.10717104.

[17] N. Rajinikanth and L. Pavithra, “Heart diseases prediction for optimization based feature selection and classification using machine learning methods,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, 2021, doi: 10.14569/ijacsa.2021.0120280.

[18] C. A. U. Hassan et al., “Effectively predicting the presence of coronary heart disease using machine learning classifiers,” Sensors, vol. 22, no. 19, 2022, doi: 10.3390/s22197227.

[19] D. Swain, S. K. Pani, and D. Swain, “A metaphoric investigation on prediction of heart disease using machine learning,” International Conference on Advanced Computation and Telecommunication (ICACAT 2018), pp. 1-6, 2018, doi: 10.1109/ICACAT.2018.8933603.

[20] K. Yuan, L. Yang, Y. Huang, and Z. Li, “Heart disease prediction algorithm based on ensemble learning,” Proceedings - 2020 7th International Conference on Dependable Systems and Their Applications, DSA 2020, pp. 293–298, 2020, doi: 10.1109/DSA51864.2020.00052.

[21] T. A. Assegie, R. Subhashni, N. K. Kumar, J. P. Manivannan, P. Duraisamy, and M. F. Engidaye, “Random forest and support vector machine-based hybrid liver disease detection,” Bulletin of Electrical Engineering and Informatics, vol. 11, no. 3, pp. 1650–1656, 2022, doi: 10.11591/eei.v11i3.3787.

[22] S. Chaitusaney and A. Yokoyama, “An appropriate distributed generation sizing considering recloser-fuse coordination,” Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference, pp. 1–6, 2005, doi: 10.1109/TDC.2005.1546838.

[23] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, “Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system,” World Journal of Engineering and Technology, vol. 6, no. 4, pp. 854–873, 2018, doi: 10.4236/wjet.2018.64057.

[24] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, “A method for improving prediction of human heart disease using machine learning algorithms,” Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/1410169.

[25] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing and Applications, vol. 29, no. 10, pp. 685–693, 2018, doi: 10.1007/s00521-016-2604-1.

[26] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari, "An empirical study on machine learning algorithms for heart disease prediction," IAES International Journal of Artificial Intelligence, vol. 11, no. 3, pp. 1066–1073, 2022, doi: 10.11591/ijai.v11.i3.pp1066-1073.