

Comparative Analysis of Heart Disease Prediction System

Kaveti Yeshwanth^{1*}, Deepika Ghai²

Abstract

In the present world, where heart illnesses are on the rise, it is crucial to forecast these diseases. Performing the task on heart disease is a bit difficult and it must be finished precisely and successfully. Heart disease identification relies heavily on Machine Learning (ML) and data mining approaches. The primary focus of the review paper is that patients are easily prone to cardiac diseases depending on medical traits. Using the patient's medical history, we compared techniques to anticipate in case the patient may be cured of heart disease or not. In this text, the editorial has practiced various types of machine learning algorithms such as Random Forest (RF), Logistic Regression (LR), Decision Trees (DT), K-Nearest Neighbor (KNN), Naive Bayes (NB), Support Vector Machine (SVM), etc. It has been observed that K- Nearest Neighbor along with Random Forest is the best among all heart disease prediction systems.

Keywords: Heart disease, Machine Learning, Data Mining, Cardiac condition, K-Nearest Neighbor, Logistic regression, Support Vector Machine, Random Forest Classifier.

INTRODUCTION

Heart disease covers a wide range of diseases that have an impact on the heart, such as infections, genetic abnormalities, and blood vessel diseases. There is a wide variety of forms included in this, and each individual has different characteristics, causes, and medication. Few people's health can be ameliorated by making a few changes in their style of living and medications [1]. On the Other side, a few of them may require an operation to bring back their performance. Cardiac disease is frequently determined by many risk factors, including vaping, more usage of alcohol and caffeine, tension, and physical inactivity, simultaneously physical attributes including adiposity, hyperconscious, excessive blood cholesterol, and pre-existing heart diseases. Taking preventative actions to avoid the consequences and the cause of this illness is largely dependent on the effective, accurate, and prior medical examination of heart disease [2].

*Author for Correspondence

Kaveti Yeshwanth
E-mail: yeshwanthchandra27@gmail.com

¹Student, School of Electronics and Electrical Engineering
Lovely, Professional University Jalandhar, India

²Professor, School of Electronics and Electrical Engineering,
Lovely Professional University Jalandhar, India

Received Date: July 21, 2023
Accepted Date: August 10, 2023
Published Date: October 12, 2023

Citation: Kaveti Yeshwanth, Deepika Ghai. Comparative Analysis of Heart Disease Prediction System. International Journal of Advance in Molecular Engineering. 2023; 1(1): 1–6p.

In the contemporary world, one of the significant threats faced in the medical field is the lack of quality service and early, accurate prediction. Automation can fix this issue with the utilization of data mining together with machine learning techniques. The hidden relationship among correlated characteristics may be found using data mining classification approaches, which is important for predicting the discrete attribute from a huge dataset. Without the assistance of medical professionals, it is simple to identify heart disease patients by examining those hidden patterns and the connected traits. The use of machine learning algorithms in heart disease prediction can aid healthcare professionals in identifying high-risk

patients and taking preventive measures to minimize the risk of heart disease. These algorithms can also provide a more personalized approach to patient care by identifying individual risk factors and customizing interventions accordingly [3].

The creation of machine learning models for heart disease prediction necessitates the collection and pre-processing of patient data from a variety of sources, including electronic health records and medical imaging data. The quality and quantity of data are critical factors that can influence the machine learning algorithm's performance [4–6]. Despite the potential advantages of machine learning algorithms in heart disease prediction, several issues must be addressed, such as data privacy and ethical considerations. The usage of confidential patient information raises concerns about data security and the possibility of patient information being misused [7, 8]. In conclusion, machine learning algorithms have the potential to enhance heart disease prediction and patient care. However, additional research is required to assess the performance of various algorithms, address data collection and privacy concerns, and ensure that patient data is used ethically. The remaining paper is organized into the following sections: Section 2 describes a literature review of heart disease. The comparative analysis of existing heart disease systems on various datasets is also included in this section. The conclusion and potential future extensions are presented in Section 3.

RELATED WORK

In recent years, many researchers have employed various machine learning and data mining approaches in a variety of studies relating to the detection of cardiac disease. Anbuselvan, P et al. [9] compared various ML techniques such as LR, NB, SVM, KNN, DT, and the ensemble technique of XGBoost (XGB). It has been found that RF and XGBoost (XGB) have the highest accuracy of 86.89% and 78.69% respectively among other ML techniques. The Cleveland heart dataset has been used in the experiment.

Hasan, S. M. M et al. [10] used info gain feature selection technique by comparing different classified techniques such that LR KNN, NB, DT (ID3), and RF. It has been observed that LR outperforms better with 92.76% by considering different classified techniques. The Cleveland dataset derived from UCI was used in this paper. They have experimented using 14 features (such that age, sex, cp, trestbps, chol, fbs, restcg, thalach, exang, oldpeak, slope, ca, thal, class) and 10 attributes (such as sex, cp, oldpeak, restcg, thalach, exang, slope, ca, thal, class). The classification accuracy has been increased when 10 attributes have been used instead of 14 attributes, thus reducing computational complexity and processing speed.

Jindal, H et al. [11] proposed a cardiovascular disease detection model using three different machine learning techniques LR, KNN, and RF Classifier. The total accuracy with LR, KNN, and RF classifiers is 87.5%. KNN and LR have achieved the best accuracy with 88.52%. They have selected the dataset from the UCI repository.

Kavitha, M et al. [12] suggested a method that is being presented as a hybrid model with high novelty that uses automated ML to treat heart disease. Three various kinds of machine-learning techniques such as DT, RF, and hybrid model (i.e., mix of both RF and DT) are used. The Cleveland dataset is used for experimentation. The accuracy with RF and Hybrid models are 81% and 88% respectively, whereas DT achieved 79% accuracy.

Ishaq, A et al. [13] proposed a model for detecting cardiovascular disease using ML techniques such as DT, LR, SVM, RF, Gradient Boosting (GBM), Extra Tree Classifier (ETC), Stochastic Gradient (SGD) Gaussian Naive Bayes (G- NB), Adaptive boosting (AdaBoost). The dataset is derived from the UCI machine learning repository. The LR classifier performed well and attained the highest accuracy of 85.56%, a precision of 85%, a recall of 86%, and an F-Score of 85% among all other ML techniques. SVM and G-NB are classified as second-best classifiers with both having the same accuracies (86.67%) and the same F-Score (86%).

Diwakar, M et al. [14] summarized and reviewed some of the techniques using machine learning and image fusion such as NB, Artificial Neural Network (ANN), KNN, and SVM. It has been concluded that to increase algorithm precision and provide accurate findings, certain feature selection techniques are applied.

Kunjachen, L. M et al. [15] used various KNN algorithms, such as the weighted K-NN, adaptive K-NN, and classical K-NN. According to the results, KNN is one of the most effective techniques with an accuracy of 97.99%. The weighted KNN model attained the second-highest value with an accuracy of 96.99%. The Kaggle dataset is used to collect the information needed to diagnose heart failure. Highly positive metrics and barely any adverse metrics are produced by the adaptive KNN model.

Akkaya, B. et al. [16] used Synthetic Minority Oversampling Technique Tomek Links (SMOTE - Tomek Link) method combining Adaboost, LR, KNN, SVM, NB, DT, Multilayer Perceptron (MLP), and XGB. According to non-outlier data, the XGB has attained an accuracy of 89%, whereas for outlier data it has achieved 84.6%. Accordingly, the KNN algorithm has gained an accuracy of 85.6% for non-outlier data, and for outlier data, it has gained 81%. XGB and K-NN algorithms have been used for initial sickness detection and pattern finding in heart disease examination.

Faieq, A. K. et al. [17] proposed two methods such as SVM and Forward Propagation Artificial Neural Network (FP-ANN). The data has been collected from the Cleveland database. Both the methods have been compared based on their sensitivity and accuracy for predicting heart disease and resulted that the SVM is the best algorithm to achieve high accuracy. SVM has obtained an accuracy of 89.1% for Arrhythmia disease, whereas FP-ANN has achieved an accuracy of 85.4%.

Abuhaija, B et al. [18] used seven different algorithms such as NB, KNN, DT, SVM, baseline accuracy, LR, and ANN multilayer perceptron. Among these classifiers, the DT (J48) has achieved the highest accuracy of 95.76% depending on a few attributes such as sensitivity, accuracy, and specificity. These are derived from Weka and Statistical Package for Social Sciences (SPSS). The data for this experiment has been gathered from the pupil in the Saudi university and lecturers of Saudi universities.

Radwan, M et al. [19] proposed a test consisting of six different ML algorithms such as KNN, Gradient Boosting, RF, NB, DT, and LR. Among these, RF algorithms have accounted for the highest accuracy of 98.6%. But after doing various tests, the LR has stayed dominant in all tests with accuracies of 91.6% and 90.8%. Three different datasets are used in this experiment with 5,74,440 entries varying between 12 to 21 features.

Dass, A. K et al. [20] analysed diverse ML techniques such as LR, DT, SVM, RF, KNN, and NB. These algorithms are compared under various factors like Recall, Precision, Area Under Curve matrices, and F1-Score to understand their efficiency. Using Jupiter notebook, they found out that the KNN has the highest precision with a precision of 91.7 % and it outperformed all other algorithms.

Researchers have used various kinds of machine-learning techniques such as LR, DT, SVM, RF, KNN, NB, etc in experimentation [21–26]. In those algorithms, a few of them has been achieved an accuracy of 97.99% which is adaptive KNN by Kunjachen, L. M et al. [15] and random forest (98.6%) by Radwan, M et al [19], and Logistic Regression (92.76%) by Hasan, S. M. M et al. [10]. So, Random Forest and K-Nearest Neighbors can achieve high accuracy. Different algorithms used for heart disease prediction is shown in Figure 1. Table 1 shows a comparative analysis of existing methods by different authors and their techniques to achieve high accuracy.

CONCLUSION AND FUTURE SCOPE

In the contemporary world, heart disease has become one of the most common diseases and a lot of people suffer and die at an early age. The cost of treating the heart patient is also increasing day by day, even though the operations are failing. So, predicting the heart will save people as well as reduce the

cost of operation. There are a wide variety of machine learning techniques present in today's world to detect heart disease. In this review, we have compared various machine learning techniques such as KNN, LR, DT, SVM, RF, etc. It has been observed that KNN and RF have the best among all other classifiers for predicting heart disease in terms of accuracy. In the future, we will try to enhance the accuracy of KNN and RF by applying hybrid techniques to improve the prediction of heart disease.

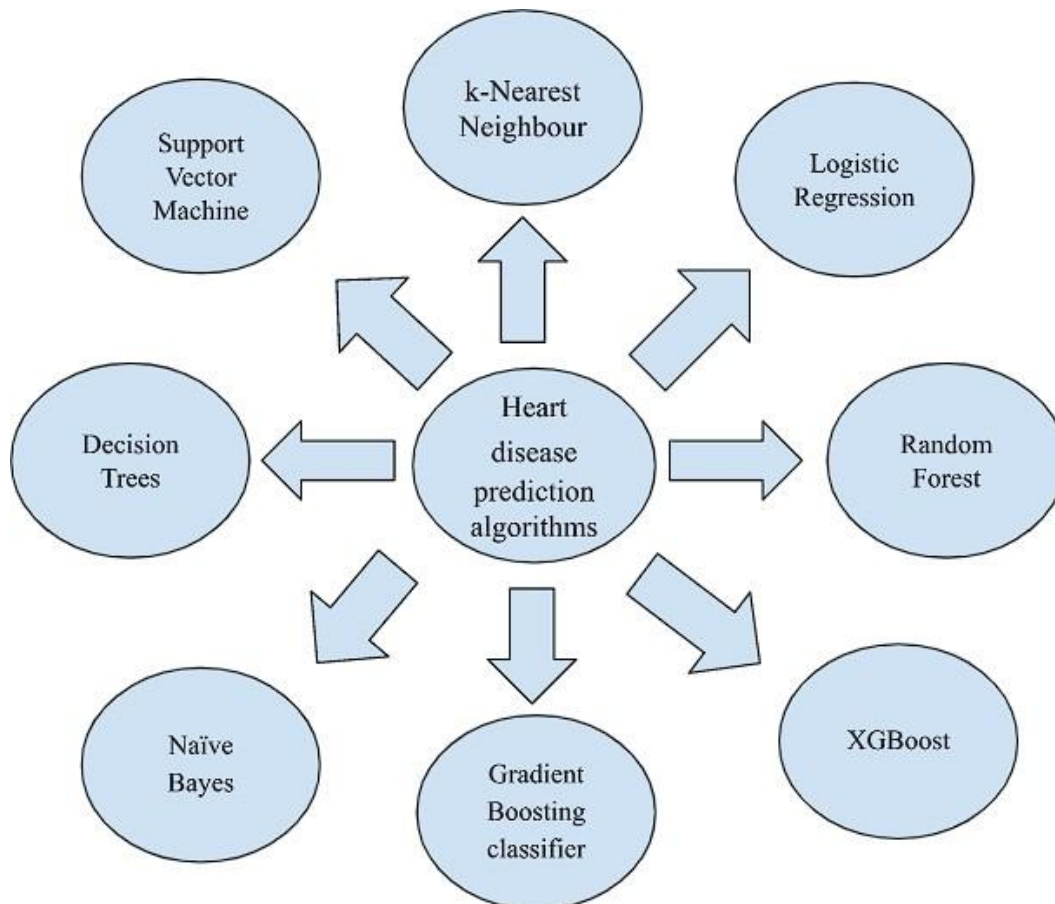


Figure 1. Different algorithms used for heart prediction.

Table 1. Comparative analysis of existing methods of heart disease prediction system.

S.N.	Author	Year	Techniques used	Accuracy
1	Anbuselvan, P et al. [9]	2020	LR, NB, SVM, KNN, DT, RF and XGB	RF (86.89%)
2	Hasan, S. M. M et al. [10]	2020	KNN, DT(ID3), G-NB, LR and RF	LR (92.76%)
3	Jindal, H et al. [11]	2020	LR, KNN and RF	KNN and LR (88.52%)
4	Kavitha, M et al. [12]	2021	RF, DT and Hybrid model	Hybrid model (88%)
5	Ishaq, A et al. [13]	2021	DT, LR, GBM, SGD, ETC, G-NB, AdaBoost and RF	LR (85.56%)
6	Kunjachen, L. M et al. [15]	2022	Weighted KNN, adaptive KNN, classical KNN	Adaptive KNN (97.99%)
7	Akkaya, B. et al. [16]	2022	LR, SVM, NB, KNN, DT, MLP and XGB	XGB (89%)
8	Faieq, A. K. et al. [17]	2022	SVM and FP-ANN	SVM (89.1%)
9	Abuhaija, B et al. [18]	2023	Baseline accuracy, NB, KNN, DT, SVM, LR, ANN and MLP	DT (J48) (95.76%)
10	Radwan, M et al [19]	2023	KNN, GBM, RF, NB, DT and LR	RF (98.6%)
11	Dass, A. K et al. [20]	2023	NB, DT, RF, KNN, SVM and LR	KNN (91.7%)

Acknowledgment

I would like to thank School of Electronics and Electrical Engineering (SEEE), Lovely Professional University for giving me this opportunity and providing all the resources for carrying out this research work.

REFERENCES

1. Rahman, B., Warnars, H. L. H. S., Sabarguna, B. S., & Budiharto, W. (2021, November). Heart Disease Classification Model Using K- Nearest Neighbor Algorithm. In 2021 Sixth International Conference on Informatics and Computing (ICIC) (pp. 1-4). IEEE.
2. Pasha, S. N., Ramesh, D., Mohmmad, S., & Harshavardhan, A. (2020, December). Cardiovascular disease prediction using deep learning techniques. In IOP conference series: materials science and engineering (Vol. 981, No. 2, p. 022006). IOP Publishing.
3. Thomas, J., & Princy, R. T. (2016, March). Human heart disease prediction system using data mining techniques. In 2016 international conference on circuit, power and computing technologies (ICCPCT) (pp. 1-5). IEEE.
4. 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.
5. Sharma, H., & Rizvi, M. A. (2017). Prediction of heart disease using machine learning algorithms: A survey. *International Journal on Recent and Innovation Trends in Computing and Communication*, 5(8), 99-104.
6. Aljanabi, M., Qutqut, M. H., & Hijjawi, M. (2018). Machine learning classification techniques for heart disease prediction: a review. *International Journal of Engineering & Technology*, 7(4), 5373-5379.
7. Rajathi, S., & Radhamani, G. (2016, March). Prediction and analysis of Rheumatic heart disease using kNN classification with ACO. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 68-73). IEEE.
8. Shanmugasundaram, G., Selvam, V. M., Saravanan, R., & Balaji, S. (2018, July). An investigation of heart disease prediction techniques. In 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA) (pp. 1-6). IEEE.
9. Anbuselvan, P. (2020). Heart disease prediction using machine learning techniques. *Int. J. Eng. Res. Technol*, 9, 515-518.
10. Hasan, S. M. M., Mamun, M. A., Uddin, M. P., & Hossain, M. A. (2018, February). Comparative analysis of classification approaches for heart disease prediction. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.
11. Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing.
12. Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1329-1333). IEEE.
13. Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. *IEEE Access*, 9, 39707-39716.
14. Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, 37, 3213- 3218.
15. Kunjachan, L. M., & Kavitha, R. (2022, December). Comparative Study on Detection of Heart Disease using KNN Algorithm. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 14-19). IEEE.
16. Akkaya, B., Sener, E., & Gursu, C. (2022, June). A Comparative Study of Heart Disease Prediction Using Machine Learning Techniques. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1- 8). IEEE.
17. Faieq, A. K., & Mijwil, M. M. (2022). Prediction of heart diseases utilizing support vector machine and artificial neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1), 374-380.
18. Abuhaija, B., Alloubani, A., Almatari, M., Jaradat, G. M., Hemn, B. A., Abualkishik, A. M., & Alsmadi, M. K. (2023). A comprehensive study of machine learning for predicting cardiovascular

- disease using Weka and SPSS tools. *International Journal of Electrical and Computer Engineering*, 13(2), 1891.
19. Radwan, M., Mohamed Abdelrahman, N., Wael Kamal, H., Khaled Abdelmonem Elewa, A., & Moataz Mohamed, A. (2023). MLHeartDisPrediction: Heart Disease Prediction using Machine Learning. *Journal of Computing and Communication*, 2(1), 50-65.
 20. Dass, A. K. (2023). Comparison of Heart Disease Prediction Using Different Machine Learning Algorithms.
 21. Shukla, P. K., Sandhu, J. K., Ahirwar, A., Ghai, D., Maheshwary, P., & Shukla, P. K. (2021). Multiobjective genetic algorithm and convolutional neural network based COVID-19 identification in chest X-ray images. *Mathematical Problems in Engineering*, 2021, 1- 9.
 22. Rani, S., Ghai, D., Kumar, S., Kantipudi, M. V. V., Alharbi, A. H., & Ullah, M. A. (2022). Efficient 3D AlexNet Architecture for Object Recognition Using Syntactic Patterns from Medical Images. *Computational Intelligence and Neuroscience*, 2022.
 23. Tripathi, S. L., Mendiratta, N., Ghai, D., Avasthi, S., & Dhir, K. (2022). Coronavirus: Diagnosis, detection, and analysis. In *Biomedical Engineering Applications for People with Disabilities and the Elderly in the COVID-19 Pandemic and Beyond* (pp. 109- 117). Academic Press.
 24. Tripathi, S. L., Dhir, K., Ghai, D., & Patil, S. (Eds.). (2021). *Health Informatics and Technological Solutions for Coronavirus (COVID- 19)*. CRC Press.
 25. Ghai, D., Tripathi, S. L., Saxena, S., Chanda, M., & Alazab, M. (2022). Machine learning algorithms for signal and image processing.
 26. Chitra, R., Warsi, K. A., Muzamil, S., Shankar, P. N., Ghai, D., & Dharmani, B. C. (2022, June). Performance Evaluation of Machine Learning Algorithms in Design and Development of Heart Disease Detection. In *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)* (pp. 1-8). IEEE.