CSVTU International Journal of Biotechnology, Bioinformatics and Biomedical

Vol. 7(3), 34-40, 2022 ISSN: 2455-5762

Heart Disease Prediction: a Machine Learning Approach

Sahu V., Sahu T., Brajesh R. G.*

Department of Biomedical Engineering and Bioinformatics
University Teaching Department, Chhattisgarh Swami Vivekanand Technical University,
Bhilai -491107, Chhattisgarh, India * Corresponding author's email:
Faculty3_bmeb@csvtu.ac.in

Received December18, 2022; received in revised form December 28, 2022; accepted December 2022; available online December 2022

Abstract

Machine learning has been one of the most widely used tools in medical science. It has shown promising application in disease detection as well as in disease prediction. In this paper we focus on the artificial intelligence-based heart disease prediction system using machine learning algorithms. We discuss various algorithms that have been employed by researchers in healthcare sector and chalk out the comparative analysis among these various algorithms. Traditional clinical diagnosis methods have their own limitations, with the help of our study we would be focusing on accuracy and ability to detect the heart disease at early stage. We have discussed some most recent works in heart disease prediction and done a comparative analysis on the accuracy and fusibility of these methods.

Keywords: Heart Disease, Machine Learning, Artificial intelligence, Machine learning Predictive analytics, Classification

Broad area: Biomedical Engineering; **Sub area:** Machine learning

Review Article

1. Introduction

World Health Organization data shows that, nearly 18 million people die because of heart related complications [1-3]. Clinically, the alcohol abuse, smoking, obesity, hypertension, hyperglycemia etc. have been identified as major causative agents for the heart diseases [4, 5]. These factors have many biochemical traceable parameters such as high cholesterol content, triglycerides, fasting plasma glucose [6] etc. Trained physicians use these factors to diagnose the heart disease conditions. However, quantitative prediction of cardiovascular diseases (CVDs) based on these inputs is still in its early stage [7]. The most common symptoms in CVDs are, pain the arm and chest area [8]. The diagnosis involves, electrocardiograph, sonography, lipid profiling and blood glucose estimation to diagnose CVDs. Unable to diagnose the early stage of CVDs is the bottleneck of these techniques [9, 10]. Cardiovascular diseases are often used to describe the condition of blockage in the heart blood vessels causing chest pain, angina, stroke and heart attack [11]. Apart from these conditions, irregular rhythm, faulty heart valve, weak heart muscles are some other kinds of heart diseases [12].

Since the genesis of machine learning, its limitless application in almost every area of science is under exploration. With recent advancement in the computation power, researchers have used complex programming in the form of deep learning in the field of healthcare such as; prediction of diseases [13], annotation of images,

identification of pattern in huge data set etc. [14].Machine learning uses algorithms to identify/classify desired data based on training from the historic data. For example, if blood glucose dataset for n individuals is given, a computer algorithm can be designed which train itself from the previously diagnosed data for identification of diabetes in unknown samples [15, 16].

The shoot up in the heart disease cases each year has motivated researchers to develop algorithms [17-19], which can predict heart related problems at early stage using existing data. Numerous supervised machine learning algorithms have been developed and their efficiency is compared in order to get accurate prediction of heart diseases. Some of the widely used algorithms are, decision tree (DT)[20], k-nearest neighbor (kNN)[21], naïve bayes (NB)[22], random forest classifier (RF)[23, 24], support vector machines (SVM)[25, 26], logistic regression (LR)[27, 28] etc. A detailed mechanism of heart disease prediction using various machine learning algorithm is discussed in figure 1 [29].

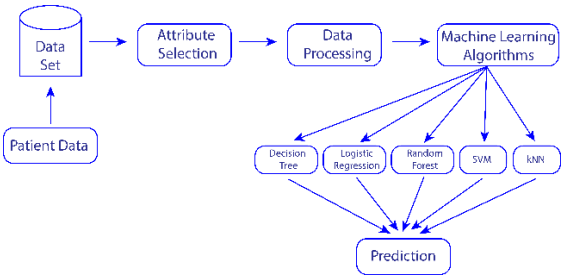

Fig 1. Illustration of disease prediction using machine learning algorithm

2. Literature Survey

This paper is written to ensemble various machine learning approach taken by researchers in order to find out the heart disease prediction. Some of the recently published articles in this domain are:

Patel et. al used different algorithms of decision tree approach and later compared their efficiency [30]. To test the output from algorithm they have used Waikato Environment for Knowledge Analysis (WEKA). The classification tree algorithm used are J48 algorithm, logistic model tree algorithm and random forest algorithm. The patient data sets used by Patel's group is taken from University of California Irvine (UCI) machine learning repository [31]. Patel's group reported that J48 algorithm is the best decision tree algorithm. J48 was found to be more accurate (55.7%) compared with logistic model and random forest algorithm [30].

Bhalerao and group used hybrid method of combined K-means and artificial neural network (ANN) algorithm. The authors have used the UCI repository for patient data. The authors claim that, the improved K-means and ANN algorithm shows 6.74% accurate percentage [32]. In future they have extended their work to accurate prediction of Cancer, HIV etc.

In 2017, Bertsimas group used patient ECG signal data from different hospitals and then applied XGBoost algorithm for heart disease prediction [33]. The prediction process can be illustrated as: signal acquisition, signal processing, feature extraction, model training finally feature prediction. The model was fast, and accuracy varied from 93-99%. Similar to the above article, Sahoo et. al in 2017 also published an article where they have used ECG signals for the heart disease prediction using hybrid system of support vector machines and neural network (NN). The accuracy achieved in their paper was 98.40%. The patient samples were retrieved from MIT database. The main focus in this study was to analyze the irregular heartbeats by accurately detecting the patterns in QRS complex [34].

Haq and group in 2018 extensively studied seven different machine learning algorithms along with different feature selection and well as classification algorithms. They have also used the UCI repository for the patient data set. Seven tested machine learning algorithms were, logistic regression, k-nearest neighbor, artificial neural network, support vector machines, naïve bayes and decision tree. To extract the features from Cleveland database they have used Relief, mRMR and LASSO tools. Overall, their study demonstrate that support vector machine have highest accuracy, specificity as well as sensitivity [35]. The detailed comparison is given in Table 1.

Table-1 Accuracy, sensitivity and specificity of different machine learning algorithms by Haq group

Predictive models	**Accuracy (%)**	**Specificity (%)**	**Sensitivity (%)**
Logistic regression	84	85	83
K-nearest neighbor	76	74	73
Artificial neural network	74	73	74
SVM	86	88	78
Naïve Bayes	83	87	78
Decision tree	74	76	68
Random forest	83	70	94

In the same year, Nashif et. al published an article showing the real time cardiovascular health monitoring system. They have also reported that number of features are correlated with the performance of the model, reduced features reduce the accuracy of the system however SVM still shows the best performance. In their work they have proposed to integrate the photoplethysmography (PPG) output to Aurdino. The server implemented with machine learning algorithm will use these real time data and help in prediction of heart disease conditions [36].

Seh et. al in 2019 wrote a review on heart disease prediction and listed some of the best known technique in machine learning algorithm such as, decision list using WEKA gives 99.62 % accuracy using UCI database [37]. Similar to the above study many authors have contributed and performed the comparative analysis between various machine learning algorithms and feature selection tools [38-42].

In 2020, Khalid's group proposed a multiple stage detection model of heart disease using five algorithms such as Random Forest, Gradient Boosting Machine, Extreme Random Forest, Logic Regression, and Support Vector Machine and compared the accuracy. They also report that; logistic regression shows the better results compared to other algorithms (82%). In future work they have proposed to work with feature selection tools to enhance the algorithm performance [43].

In 2022, Ahmed et. al, reported a self-augmented datasets of heart patients using multiple machine learning models. In this study, they have come up with hybrid machine learning algorithms to predict heart failure conditions. The ensemble algorithm and their accuracy are listed below. Gaussian Naïve Bayes (GNB – 67%), Gradient Boosting Classifiers (GBC – 88%), Multinomial Naïve Bayes (MNB – 85%), CatBoost (89%), XGBoost (84%). In future this research can be upgraded to predict the survival of patients by using HF dataset [8].

In 2022, Absar et. al used four machine learning algorithms (RF, DT, AB, and

KNN) to predict the coronary heart diseases. They have used patient dataset using CHSLB (Cleveland, Hungary, Switzerland, and Long Beach) and Cleveland datasets. The extended dataset has given the 100% prediction using KNN algorithm on the other hand it gives 97.8% accuracy using Cleveland dataset. The CHSLB dataset shows accuracy of 99.025%, 96.103%, and 100%, in RF, AB, and DT models respectively. All these algorithms have demonstrated the powerful technique of machine learning in heart disease prediction.

3. Discussion

Diagnosis of cardiac disease is one of the most challenging job in medical science. Critical examination by medical practitioner requires a huge investment of money skill etc. At the end there will be always a chance of human error. With advancement of technology we can certainly use the computation powers in the field of medical science to predict the likelihood of certain disease based on previously stored data. Machine learning is one such tool to provide highly accurate and less time-consuming platform. In this paper such tools which have been previously reported by researchers are compiled and summarized.

We report that stand-alone algorithms have limitations, we need more ensembled algorithms to rectify these limitations and increase the accuracy, specificity and sensitivity of our model.

Conflict of interest

There is no conflict of interest among the authors.

Author contribution

Authors share equal contribution, VS has done the literature survey, TS have prepared the manuscript and RG have compiled and edited the manuscript.

Acknowledgement

Authors are grateful towards department of Biomedical engineering for its resources.

Reference

[1] Mehra R 2007 Global public health problem of sudden cardiac death *Journal of electrocardiology* **40** S118-S22

[2] Okrainec K, Banerjee D K and Eisenberg M J 2004 Coronary artery disease in the developing world *American heart journal* **148** 7-15

[3] Rayner M, Allender S, Scarborough P and Group B H F H P R 2009 Cardiovascular disease in Europe *European Journal of Cardiovascular Prevention & Rehabilitation* **16** S43-S7

[4] Hollan I, Dessein P, Ronda N, Wasko M, Svenungsson E, Agewall S, Cohen-Tervaert J, Maki-Petaja K, Grundtvig M and Karpouzas G 2015 Prevention of cardiovascular disease in rheumatoid arthritis *Autoimmunity reviews* **14** 952-69

[5] Huang S, Li J, Shearer G C, Lichtenstein A H, Zheng X, Wu Y, Jin C, Wu S and Gao X 2017 Longitudinal study of alcohol consumption and HDL concentrations: a community-based study *The American journal of clinical nutrition* **105** 905-12

[6] Vasheghani-Farahani A, Nouri N, Seifirad S, Sheikh Fathollahi M, Hakki E, Alidoosti M, Davoodi G, Masoudkabir F and Poorhosseini H 2013 Comparison of cardiovascular risk factors and biochemical profile in patients with cardiac syndrome X

and obstructive coronary artery disease: A propensity score-matched study *ARYA atherosclerosis* **9** 269-73

[7] Sergi G, Veronese N, Fontana L, De Rui M, Bolzetta F, Zambon S, Corti M-C, Baggio G, Toffanello E D and Crepaldi G 2015 Pre-frailty and risk of cardiovascular disease in elderly men and women: the Pro. VA study *Journal of The American college of cardiology* **65** 976-83

[8] Ahmed S, Shaikh S, Ikram F, Fayaz M, Alwageed H S, Khan F and Jaskani F H 2022 Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models *Journal of Sensors* **2022**

[9] Ma J-L, Cui Y-L and Dong M-C 2013 An effective low-complexity multi-vital-signs compression technique for embedded-link e-home healthcare. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*: IEEE) pp 1177-81

[10] Zeinalnezhad M, Chofreh A G, Goni F A, Klemeš J J and Sari E 2020 Simulation and improvement of patients' workflow in heart clinics during COVID-19 pandemic using timed coloured petri nets *International Journal of Environmental Research and Public Health* **17** 8577

[11] Chang V, Bhavani V R, Xu A Q and Hossain M 2022 An artificial intelligence model for heart disease detection using machine learning algorithms *Healthcare Analytics* **2** 100016

[12] Learning M 2017 Heart disease diagnosis and prediction using machine learning and data mining techniques: a review *Advances in Computational Sciences and Technology* **10** 2137-59

[13] Priya A, Garg S and Tigga N P 2020 Predicting anxiety, depression and stress in modern life using machine learning algorithms *Procedia Computer Science* **167** 1258-67

[14] Shailaja K, Seetharamulu B and Jabbar M 2018 Machine learning in healthcare: A review. In: *2018 Second international conference on electronics, communication and aerospace technology (ICECA)*: IEEE) pp 910-4

[15] Khan F A, Zeb K, Al-Rakhami M, Derhab A and Bukhari S A C 2021 Detection and prediction of diabetes using data mining: a comprehensive review *IEEE Access* **9** 43711-35

[16] Li M and Zhou Z-H 2007 Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **37** 1088-98

[17] Barrett M, Boyne J, Brandts J, Rocca B-L, De Maesschalck L, De Wit K, Dixon L, Eurlings C, Fitzsimons D and Golubnitschaja O 2019 Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care *Epma Journal* **10** 445-64

[18] Paradis V, Cossette S, Frasure-Smith N, Heppell S and Guertin M-C 2010 The efficacy of a motivational nursing intervention based on the stages of change on self-care in heart failure patients *Journal of Cardiovascular Nursing* **25** 130-41

[19] Olsen C R, Mentz R J, Anstrom K J, Page D and Patel P A 2020 Clinical

applications of machine learning in the diagnosis, classification, and prediction of heart failure *American Heart Journal* **229** 1-17

[20] Shouman M, Turner T and Stocker R 2011 Using decision tree for diagnosing heart disease patients. In: *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*, pp 23-30

[21] Deekshatulu B and Chandra P 2013 Classification of heart disease using k-nearest neighbor and genetic algorithm *Procedia technology* **10** 85-94

[22] Gupta A, Kumar L, Jain R and Nagrath P 2020 Heart disease prediction using classification (naive bayes). In: *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*: Springer) pp 561-73

[23] Pal M and Parija S 2021 Prediction of heart diseases using random forest. In: *Journal of Physics: Conference Series*: IOP Publishing) p 012009

[24] Dhanka S and Maini S 2021 Random Forest for Heart Disease Detection: A Classification Approach. In: *2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES)*: IEEE) pp 1-3

[25] Shah S M S, Shah F A, Hussain S A and Batool S 2020 Support vector machines-based heart disease diagnosis using feature subset, wrapping selection and extraction methods *Computers & Electrical Engineering* **84** 106628

[26] Son Y-J, Kim H-G, Kim E-H, Choi S and Lee S-K 2010 Application of support vector machine for prediction of medication adherence in heart failure patients *Healthcare informatics research* **16** 253-9

[27] Ambrish G, Ganesh B, Ganesh A, Srinivas C and Mensinkal K 2022 Logistic Regression Technique for Prediction of Cardiovascular Disease *Global Transitions Proceedings*

[28] Zhang Y, Diao L and Ma L 2021 Logistic Regression Models in Predicting Heart Disease. In: *Journal of Physics: Conference Series*: IOP Publishing) p 012024

[29] Sireesha V, Hegde N P, Nanditha P and Naresh B R 2022 *Smart Intelligent Computing and Applications, Volume 2*: Springer) pp 181-90

[30] Patel J, TejalUpadhyay D and Patel S 2015 Heart disease prediction using machine learning and data mining technique *Heart Disease* **7** 129-37

[31] Lichman M 2013 UCI machine learning repository. Irvine, CA, USA)

[32] Bhalerao S and Gunjal B 2013 Hybridization of improved k-means and artificial neural network for heart disease prediction *Int. J. Comput. Sci. Trends Technol* **4** 5461

[33] Bertsimas D, Mingardi L and Stellato B 2021 Machine learning for real-time heart disease prediction *IEEE Journal of Biomedical and Health Informatics* **25** 3627-37

[34] Sahoo S, Kanungo B, Behera S and Sabut S 2017 Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities *Measurement* **108** 55-66

[35] Haq A U, Li J P, Memon M H, Nazir S and Sun R 2018 A hybrid intelligent system framework for the prediction of heart disease using

machine learning algorithms *Mobile Information Systems* **2018**

[36] Nashif S, Raihan M R, Islam M R and Imam M H 2018 Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system *World Journal of Engineering and Technology* **6** 854-73

[37] Seh A H and Chaurasia P K 2019 A review on heart disease prediction using machine learning techniques *Journal Homepage: http://www. ijmra. us* **9**

[38] Ramalingam V, Dandapath A and Raja M K 2018 Heart disease prediction using machine learning techniques: a survey *International Journal of Engineering & Technology* **7** 684-7

[39] Sujatha P and Mahalakshmi K 2020 Performance evaluation of supervised machine learning algorithms in prediction of heart disease. In: *2020 IEEE International Conference for Innovation in Technology (INOCON)*: IEEE) pp 1-7

[40] Al'Aref S J, Anchouche K, Singh G, Slomka P J, Kolli K K, Kumar A, Pandey M, Maliakal G, Van Rosendael A R and Beecy A N 2019 Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging *European heart journal* **40** 1975-86

[41] Ali M M, Paul B K, Ahmed K, Bui F M, Quinn J M and Moni M A 2021 Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison *Computers in Biology and Medicine* **136** 104672

[42] Rani P, Kumar R, Ahmed N M and Jain A 2021 A decision support system for heart disease prediction based upon machine learning *Journal of Reliable Intelligent Environments* **7** 263-75

[43] Amen K, Zohdy M and Mahmoud M 2020 Machine learning for multiple stage heart disease prediction. In: *Proceedings of the 7th International Conference on Computer Science, Engineering and Information Technology*, pp 205-23