International Journal on Advances in Engineering Technology and Science
Volume: 5, Issue: 1, 2024, e-ISSN: 2455-3131
DOI: 10.5281/zenodo.10718879

A Review of Recent Studies on Prediction of Cardiovascular Disease

Irfan Khan[1], Pinaki Ghosh[2]

[1]PhD Scholar, [2]Professor
Sanjeev Agrawal Global Educational University, Bhopal

[1]irfank92@gmail.com, [2]pinaki.g@sageuniversity.edu.in

***Abstract* - The heart is a vital organ of the human body. It's the main part of our circulation system, and cardiovascular disease has been a common cause of mortality in the last few decades. It's increasing day by day at a rapid rate. So, it is necessary to build a system to diagnose cardiovascular disease beforehand. Machine learning is a branch of artificial intelligence; it learns from historical data, builds prediction models, and, whenever it receives new input data, predicts the outcome. The authors discussed the various machine learning algorithms used to measure the accuracy of cardiovascular disease. The prime contribution of our work is to study the various machine learning techniques used to measure accuracy to predict heart disease.**

Keywords: **cardiovascular disease, machine learning, supervised, unsupervised. Logistic regression.**

I. INTRODUCTİON

Cardiovascular disease is used to describe a group of conditions that affect the blood vessels and heart. In CD blood vessels, coronary artery disease, and issues with the heart muscle (cardiomyopathy), diseases may occur. Other problems in heart disease include congenital heart defects, arrhythmias, and heart valve disease [1].

Cardiovascular disease is a serious and common health problem that can result in various symptoms, including heart failure, heart attack, stroke, and even mortality. Medical conditions such as high blood pressure, high cholesterol, diabetes, and obesity are the reasons for cardiovascular disease, and various risk factors are also included, such as an unhealthy diet, smoking, and a lack of physical activity or exercise. Heart disease can also be dependent on age, sex, and family history [1].

To prevent and control the conditions of heart disease, a healthy lifestyle and the management of risk factors are very necessary. This involves regular physical activities, a good diet plan, maintaining weight, quitting smoking, and managing chronic medical conditions.

Prevention is good, but the heart is the most important and significant body organ; there is also a need for a system to diagnose diseases beforehand. Due to instrument inaccuracy, the majority of patients die as a result of diseases that are discovered too late. So, machine learning techniques can build a system to predict cardiovascular diseases [16, 17]. ML is an effective technology for dataset training and testing. Machine learning is a system that learns from data and experience, builds predictive models, and, whenever it receives new input data, predicts the outcome for it [2].

II. MACHİNE LEARNİNG TECHNİQUES

A machine learning approach learns from historical data, builds prediction models, and, whenever it receives new input data, predicts the outcome for it. The accuracy of predicted output relies on the quantity of data, as the massive quantity of data allows the construction of a more useful model that predicts the outcome more accurately. Earning, testing and training are the two concepts that make it more effective. On the basis of training or learning, the model learns from data and experience, and then, with the use of the required machine learning algorithm, tests will apply to different kinds of needs. Generally, machine learning is classified into three types [2].

A. Supervised Learning

Supervised learning is described as learning with proper guidance or labeled data always available at the time of learning. This learning is based on the "training me" concept. For the prediction of given data, labeled data is used as a guide. The labeled data that shows some input data is already available with the correct results. The labeled data is always present while testing new data. [2] The following processes are used in supervised learning:

- Regression
- Classification

B. Unsupervised Learning

Unsupervised learning is also a type of machine learning that is described as learning without any guidance or labeled data available. Unsupervised learning autonomously analyzes data to identify patterns and relationships between patterns.

According to these relationships, when new data is given, it classifies and stores the data in one of those relationships. That's unsupervised learning, also referred to as "self-sufficient". [2] Unsupervised algorithms go through the following methods:

- Clustering
- Associative

III. LİTERATURE REVİEW

Many researchers have worked on the prediction of cardiovascular disease due to the large number of deaths in the last few decades. To detect cardiac disease, it becomes necessary to deal with heart-related issues. Many domains are contributing to this work, including data mining, artificial intelligence, machine learning, and deep learning.

Sakthivel, M. et al. [3] proposed research on deep learning and auto-encoders (DAE) to detect human cardiac arrest on their own. Auto-encoders are employed to detect cardiovascular abnormalities. The results show that DAE has a noteworthy 93% accuracy rate in identifying human cardiac arrest, underscoring its potential for automated detection and offering a promising avenue for future investigation. The algorithm is compared to other existing approaches.

Javeed et al. [4] developed a floating window feature selection method (FWAFE) using the Cleveland database dataset and ANN and DNN classification techniques. The effectiveness of these techniques was evaluated using metrics like accuracy, specificity, sensitivity, MCC, and ROC. The DNN system performed better, with an accuracy of approximately 93.3% compared to the ANN system's 91.1%.

SM Nagarajan et al. [5] proposed a hybrid model that uses a genetic-based crow search algorithm (GCSA) for feature selection and deep convolutional neural networks (DCNN) for feature classification. In comparison to the other feature selection techniques, the suggested model, GCSA-DCNN, achieves a classification accuracy of over 91.78%, indicating an improvement in classification accuracy.

El-Shafiey, Mohamed G, et al. [6] proposed GAPSO-RF, a hybrid genetic algorithm and particle swarm optimization optimized approach based on random forest, which enhances heart-disease prediction accuracy. It uses multivariate statistical analysis, a discriminate mutation strategy, and a modified GA for global search and PSO for local search, achieving rehabilitation of individuals rejected in selection.

Jafar Abdollahi et al. [7] developed an ensemble classification model and a genetic algorithm for feature selection, achieving an accuracy of 97.57% on datasets. This model, which combined a genetic algorithm and ensemble classification, is more accurate than previous methods and suitable for healthcare implementation in identifying heart disease.

Nandy, Sudarshan, et al. [8] Introduced an intelligent healthcare framework that employs a Swarm-Artificial Neural Network (Swarm-ANN) approach to predict cardiovascular heart disease. The framework creates predetermined neural networks, adjusts their weights, and shares the globally optimized weight with other neurons. This Swarm-ANN strategy attains an accuracy of 95.78%, surpassing conventional learning methods

R. Rajendran et al. [9] developed a machine learning technique that uses entropy-based feature engineering and pre-processing to predict heart disease accurately. The technique uses a heart disease dataset from various databases and IMV + OR pre-processing. The experimental results showed improved performance in NB and LR classifiers, with an ensemble model outperforming state-of-the-art results.

Ali et al. [10] used the Kaggle dataset to predict diseases using interquartile range and synthetic minority oversampling techniques. They applied six machine learning algorithms: decision tree, random forest, logistic regression, AdaboostM1, multilayer perceptron, and k- nearest neighbours, and compared their accuracy, sensitivity, and specificity using a confusion matrix.

BP Doppala et al. [11] proposed a hybrid approach that combines radial basis functions (RBF) and genetic algorithms (GA) to more accurately detect cardiovascular disease. After reducing the number of attributes, the system was able to predict with 85.40% accuracy when using 14 attributes and 94.20% accuracy when using nine attributes.

P. Rani et al. [12] developed a system for early heart disease detection using clinical parameters, including multivariate imputation, hybridized feature selection algorithm, SMOTE, and standard scalar methods. The system, tested on the Cleveland heart disease dataset, achieved an accuracy of 86.6%, surpassing existing prediction systems.

A. Singh [13] also used the heat diseases UCI repository dataset and comparison on the basis of the accuracy of the decision tree, linear regression, KNN, and KNN on the basis of a confusion matrix. And KNN has the best accuracy of 87% of algorithms.

Shah et al. [14] proposed KNN algorithm that performed effectively with an accuracy of 90.78 amongst the decision tree, random forest, and naive bayes algorithms and used the heart disease UCI repository dataset, which has 303 instances within the 14 essential attributes of 76 attributes, including age, Sex, Cp, Fbs, cholesterol, etc.

In this literature review, many researchers have used a small dataset of 303 instances. It is suggested that in the future, larger datasets with more attributes be explored, as well as the fusion of distinct datasets to improve the diagnostic process using advanced dimension reductions and deep learning techniques.

TABLE-1 Heart Disease UCI Repository

Sr. No.	Attributes	Description
1	Age	Age of Patient's (29-58)
2	Sex	Sex (female-0, male-1)
3	Cp	Chest Pain types "1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic"
4	chol	Serum Cholesterol
5	Trestbps	Resting Blood Pressure - in mmHg
6	Restecg	Resting Electrocardiographic result (0 to 1)
7	Fbs	Fasting Blood Sugar
8	Exang	Exercise-induced angina (1-yes, 0-no)
9	Thalch	Maximums Heart Rate
10	Oldpeak	ST depression
11	Slope	ST segment (the slope of the peak exercise)
12	Ca	Fluoroscopy colored number of major vessels "0 -3"
13	Thal	Thalassemia
14	Targets	0 for no-disease & 1 for disease

IV. COMPARİSON OF WORK

Researchers have used the different datasets are Cleveland, Hungary, Switzerland, and Long Beach V downloaded from UCI and Kaggle [15], All the dataset have common 14 essential attributes of 76 attributes, including age, sex, Cp, Fbs, cholesterol, etc., as shown in Table 1. They have worked feature engineering techniques, classification techniques and hybrid approaches to acquire effective, correct, and accurate results for their proposed work with the use of multiple constraints, a confusion matrix, and an evaluation matrix. In this work, we have reviewed the comparison study of their work for the diagnosis of heart disease, shown in Table 2.

V. CONCLUSİON AND FUTURE SCOPE

In all living things, the heart is one of the most important organs. The ratio of cardiac deaths is increasing rapidly day by day, so we need a system to predict with the greatest accuracy, correctness and effectiveness to overcome this problem. In this work, we study the research work of many researchers, and the analysis shows various technologies and different attributes are used to detect the disease.

In some research work, it is shown that decision trees, random forest algorithms with an accuracy of 100%, and the KNN algorithm also perform well with an accuracy of 90.1% when using the heart disease UCI repository dataset. Therefore, the accuracy of various technologies depends on the number of attributes employed and the tool used for implementation. With this study, more complicated and combinational models still need to be used to achieve higher accuracy for heart disease early prediction.

In this study, researchers have mostly used UCI dataset have 303 instances, so there is requirement of large amount of dataset for further study and researchers can also work on native feature engineering, feature fusion, feature selection and dimension reduction methods to diagnose cardiovascular disease with great accuracy.

TABLE- 2. A Comparative Study of Literature

S. No.	Authors	Dataset Used	Feature Engineering Techniques	Classification Techniques	Accuracy
1	Shah et al (2020)	University of California, Irvine (UCI), 303 Instances	N/A	Naïve Bayes KNN Decision tree Random forest	88.16% 90.79% 80.26% 86.84%
2	A Singh et al (2020)	University of California, Irvine (UCI), 303 Instances	N/A	SVM Linear Regression Decision Tree KNN	83% 78% 79% 87%
3	Javeed et al. (2020)	University of California, Irvine (UCI), 303 instances	FWAFE (floating window for adaptable size)	ANN DNN	91.1 93.3
4	SM Nagarajan et al. (2021)	University of California, Irvine (UCI), 303 instances	genetic-based crow search algorithm (GCSA)	DCNN (Deep Convolutional Neural Network)	91.78%
5	P. Rani et al. (2021)	University of California, Irvine (UCI), 303 instances	Genetic Algorithm (GA)	SVM Naïve Bayes Logistic Regression Random forest Adaboost	86.6%
6	BP Doppala et al (2021)	University of California, Irvine (UCI), 303 instances	Genetic Algorithm (GA)	RBF (Radial basis functions) Network	94%
7	Ali, Md Mamun, et al (2021)	Hungary, Cleveland, Switzerland, and Long Beach V, Kaggle, 1025 instances	N/A	Logistic Regression AdaboostM1 MLP KNN Decision tree Random forest	89.63% 95.02% 97.95% 100.0% 100.0% 100.0%
8	El-Shafiey Mohamed G, et al. (2022)	I. University of California II. Cleveland and Statlog	Genetic Algorithm and Particle Swarm Optimization (GAPSO)	Random forest	95.6%
9	Jafar Abdollahi et al (2022)	University of California, Irvine (UCI), 270 instances	Genetic Algorithm (GA)	Ensemble Algorithm (SVM, NB, DT, MLP, KNN, RF & LR)	97.57%
10	R. Rajendran et al (2022)	Cleveland, V A medical center, Hungarian and Switzerland databases, 920 instances	Imputing missing values (IMV) Outliers are removed (OR) Entropy based FE	Ensemble Algorithm (NB & LR)	96.8%
11	Sakthivel, M. et al. (2023)	Kaggle Framingham heart dataset, 4238 instances	Sparse Auto-encoder (SAE)	(DAE) Deep learning and auto-encoders	93%
12	Nandy, Sudarshan, et al. (2023)	University of California, Irvine (UCI), 303 instances	Swarm optimization	ANN	95.78

REFERENCES

[1] J.M. Rippe, "Lifestyle strategies for risk factor reduction, prevention, and treatment of cardiovascular disease," Am. J. Lifestyle Med. 13 (2), 204–212 2018.

[2] Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. "Introduction to machine learning for brain imaging." Neuroimage, 56(2), 387-399 2011.

[3] Sakthivel M, SivaSubramanian S, Prasad GN, Thangamani M. "Automated detection of cardiac arrest in human beings using auto encoders." Measurement: Sensors. June 1; 27:100792 2023.

[4] Javeed A, Rizvi SS, Zhou S, Riaz R, Khan SU, Kwon SJ. "Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification." Mobile Information Systems. Aug 26; 2020:1-1 2020.

[5] Nagarajan SM, Muthukumaran V, Murugesan R, Joseph RB, Meram M, Prathik A. "Innovative feature selection and classification model for heart disease prediction." Journal of Reliable Intelligent Environments. Dec 8(4):333-43 2022.

[6] M.G. El-Shafiey, A. Hagag, E.S.A. El-Dahshan, M.A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," Multimed. Tool. Appl. 81 (13) 18155–18179 2022.

[7] Abdollahi J, Nouri-Moghaddam B. "A hybrid method for heart disease diagnosis utilizing feature selection-based ensemble classifier model generation." Iran Journal of Computer Science. Sep 5(3):229-46 2022.

[8] Nandy S, Adhikari M, Balasubramanian V, Menon VG, Li X, Zakarya M. "An intelligent heart disease prediction system based on swarm-artificial neural network." Neural Computing and Applications. July 35(20):14723-37 2023.

[9] Rajendran R, Karthi A. "heart disease prediction using entropy- based feature engineering and ensembling of machine learning classifiers." Expert Systems with Applications. Nov 30; 207:117882 2022.

[10] Ali, Md Mamun, et al. "heart disease prediction using supervised machine learning algorithms: performance analysis and comparison." Computers in Biology and Medicine 136 104672 2021.

[11] Doppala BP, Bhattacharyya D, Chakkravarthy M, Kim TH. "A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset." Distributed and Parallel Databases. Mar 1-20 2021.

[12] Rani P, Kumar R, Ahmed NM, Jain A. "A decision support system for heart disease prediction based upon machine learning." Journal of Reliable Intelligent Environments. Sep 7(3):263-75 2021.

[13] Singh, Archana, and Rakesh Kumar. "Heart disease prediction using machine learning algorithms." 2020 international conference on Electrical and electronics engineering (ICE3). IEEE, 2020.

[14] Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." SN Computer Science 1.6 2020.

[15] Katarya, Rahul, and Sunit Kumar Meena. "Machine learning techniques for heart disease prediction: a comparative study and analysis." Health and Technology 11.1 87-97 2021.

[16] Ghosh, Pinaki, Umesh Kumar Lilhore, Sarita Simaiya, Atul Garg, Devendra Prasad, and Ajay Kumar. "Prediction of the risk of heart attack using Machine Learning Techniques." In Data, Engineering and Applications: Select Proceedings of IDEA 2021, pp. 613-621. Singapore: Springer Nature Singapore, 2022.

[17] Ghosh, Pinaki, Devendra Prasad, and Kalpna Guleria. "An m-IoT Framework for Remote Monitoring of ECG Signals." Journal of Advanced Research in Dynamical and Control Systems 12, no. 8 (2020): 296-300.

