iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37202

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Review on Machine Learning Techniques for Diagnosis of Heart Disease

Harshal P. Sabale[1], Prof. Ravi V. Mante[2]
[1]*M.Tech.*, [2]*Assistant Professor, Department of Computer Science and Engineering, Government College of Engineering, Amravati*

Abstract: ***Now-a-days, heart disease is becoming a concern to human health. According to World Health organisation (WHO), heart disease is the number one killer among other fatal diseases. Excessive smoking, alcohol consumption and junk food are culprit for the heart disease. Physical inactivity is also a concerning to the human health. Heart disease is pretty hard to predict or diagnose using traditional methods like counselling. But, now-a-days, medical fields are using machine learning to predict or diagnose different diseases. Implementation of machine learning techniques provides faster and mostly accurate results. This can save many life. In this paper, different machine learning approach for heart disease diagnosis are reviewed.***
Keywords: Heart disease, CVD, Machine Learning

I. INTRODUCTION

Human beings are advancing very fast. This requires hard work and dedication toward it. But, due to this, human being are not taking health seriously. Unhealthy diet, smoking, etc. are key factors in affecting the health. Heart disease is one of the most common disease. Heart disease is also known as cardiovascular disease. According to World Health Organization (WHO), cardiovascular disease (CVD) or heart disease is the number one cause of death globally [1]. CVD is a group of disorders of the heart and blood vessels. CVD includes, Coronary heart disease, Cerebrovascular disease, Peripheral arterial disease, Rheumatic heart disease, Congenital heart disease, Deep vein thrombosis and pulmonary embolism. Heart disease is mainly caused by a blockage that prevents blood from flowing to the heart or brain. The most common reason for this is a build-up of fatty deposits on the inner walls of the blood vessels that supply the heart or brain.[1] Heart disease have some traditional risk factors, such as smoking, unhealthy diet and obesity, hypertension, diabetes, physical inactivity, harmful use of alcohol and hyperlipidaemia.[1] The symptoms of cardiovascular disease is chest pain, breathlessness, swollen limbs, etc. Diagnosis of heart disease is traditionally done by the analysis of the medical history of the patient, physical examination report and analysis of concerned symptoms by a physician. But, the results obtained from this diagnosis method are not accurate in identifying the patient of heart disease. Moreover, it is costly and computationally hard to analyse.[2] The diagnosis of heart disease can be easier if modern technology are used. Machine learning is being used by many researchers to provide an efficient and accurate way to detect the heart disease. Cleveland and Statlog dataset are the most popular dataset among the researchers. Both the dataset are publically available on the UCI machine learning repository. Cleveland dataset contains 76 attributes. 14 attributes are considered as important to make a classification. These same attributes are available in Statlog dataset. These attributes and their description are as shown in the below table.

TABLE I Attribute description of dataset

Sr. No.	Attribute	Description
1	age	age in years
2	sex	sex (1 = male; 0 = female)
3	cp	chest pain type
4	trestbps	resting blood pressure (in mm Hg)
5	chol	serum cholestoral (in mg/dl)
6	fbs	fasting blood sugar > 120 mg/dl
7	restecg	resting electrocardiographic results
8	thalach	maximum heart rate achieved
9	exang	exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	the slope of the peak exercise ST segment
12	ca	number of major vessels (0-3) colored by flourosopy
13	thal	Thallium Scan
14	num	the predicted attribute

The Cleveland dataset is clinically acquired datasets.[15] That is one of the main reason of popularity among the researchers. These dataset are used to train and test the machine learning model.

II. LITERATURE REVIEW

Jian Ping Li, et. el.[2] proposed machine learning based diagnosis system for heart disease. In this system, many machine learning algorithms are tested in the system. Machine learning classifiers include logistic regression, k – nearest neighbour, artificial neural network, support vector machine, naïve bayesian and decision tree. Four standard feature selection algorithms including Relief, MRMR, LASSO, LLBFS, and proposed a novel feature selection algorithm FCMIM used to solve feature selection problem. FCMIM and support vector machine (FCMIM-SVM) algorithm performed the best among other algorithms and achieved the accuracy of 92.37%. Norma Latif Fitriyani, et. el.[3] suggested a heart disease prediction model. In this methodology, there are mainly three different step. Firstly, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to detect and eliminate the outliers. Then, a hybrid Synthetic Minority Over-sampling Technique- Edited Nearest Neighbor (SMOTE-ENN) is used to balance the dataset. And, XGBoost is used to predict the heart disease. In this methodology, two dataset are used to train and test the model which are Statlog and Cleaveland. The accuracy of the model with the Statlog dataset is 95.90%. The accuracy of the model is 98.40% for Cleaveland datasets. In this method, the missing data is ignored and removed from the dataset.

Senthilkumar Mohan, et. el.[4] proposed a model which is build using combination of Random Forest and Linear Method. This model is referred as HRFLM by the author. HRFLM is trained and tested using Cleaveland dataset. This method achieved an accuracy of 88.7% which was quiet good. Ankur Gupta et. el.[5] developed a machine learning model consisting of factor analysis of mixed data (FAMD) and Random Forest-based machine learning algorithm. Factor analysis of mixed data (FAMD) is used to find the relevant features. Random Forest is used as a classifier. The cleaveland dataset is used in this methodology. The missing value is imputed using majority label. The experimental results showed that the proposed model achieved accuracy of up to 93.44%.

Ashir Javeed, et. el.[6] proposed a diagnostic model which uses random search algorithm (RSA) for features selection. Random forest model is used as classier algorithm. In this model, cleaveland dataset is used to train and test the model. By using random search algorithm, only 7 features are used to train and test the model. This method achieved 3.3% higher accuracy than conventional random forest model. This method outperformed and achieved the 93.33% accuracy. It also improves the training accuracy of the conventional random forest model.

Liaqat Ali et. al.[7] proposed a hybrid grid search algorithm (HGSA). This model is capable of optimizing the L1 regularized support vector machine and L2 regularized support vector machine simultaneously. L1 regularized support vector machine is used to remove the irrelevant feature. This is done by shrinking their coefficient to zero. While, L2 regularized support vector machine is used as a predicative model. This model achieved the maximum accuracy of 91.83%. Haq et. al.[8] performed a comparative study on a hybrid model based on various feature selection techniques such as relief, minimal-redundancy maximal-relevance (mRMR), least absolute shrinkage and selection operator (LASSO) and machine learning models such as logistic regression (LR), K-nearest neighbour (k-nn), artificial neural network (ANN), support vector machine (SVM), decision tree (DT), naïve bayes (NB) and random forest (RF). Combination of Relief-based feature selection and Logistic regression achieved accuracy up to 89%.

Shah et. al.[9] developed a selection strategy using the probabilistic principal component analysis (PPCA). PPCA is used to obtain the most significant characteristics for heart disease prediction. In this study, three dataset are used which are Cleveland, Hungarian and Switzerland dataset. This technique achieved an accuracy of 82.18%, 85.82% and 91.30% for Cleveland, Hungarian and Switzerland dataset respectively.

Latha and Jeeva[10] proved that the ensemble techniques such as bagging and boosting, are effective in improving the accuracy of weak classifier. With the help of ensemble classification, increase of 7% accuracy is achieved for weak classifiers. In this methodology, many machine learning algorithm are used which are support vector machine, naïve bayes, bayes net, c4.5, multilayer perceptron, random forest and PART. Naïve bayes, bayes net, random forest and multilayer perceptron together performed best among others. This model achieved accuracy of 85.48%. Nahato et. el.[11] proposed heart disease classifier model based on chaos firefly algorithm and rough sets-based attribute reduction (CFARS-AR). To reduce the attributes, rough sets is used. And, the chaos firefly algorithm was used to classify the disease. This model performed better as compared to naïve bayes (NB), support vector machine (SVM) and artificial neural network (ANN). This model achieved the accuracy of 88.3%.

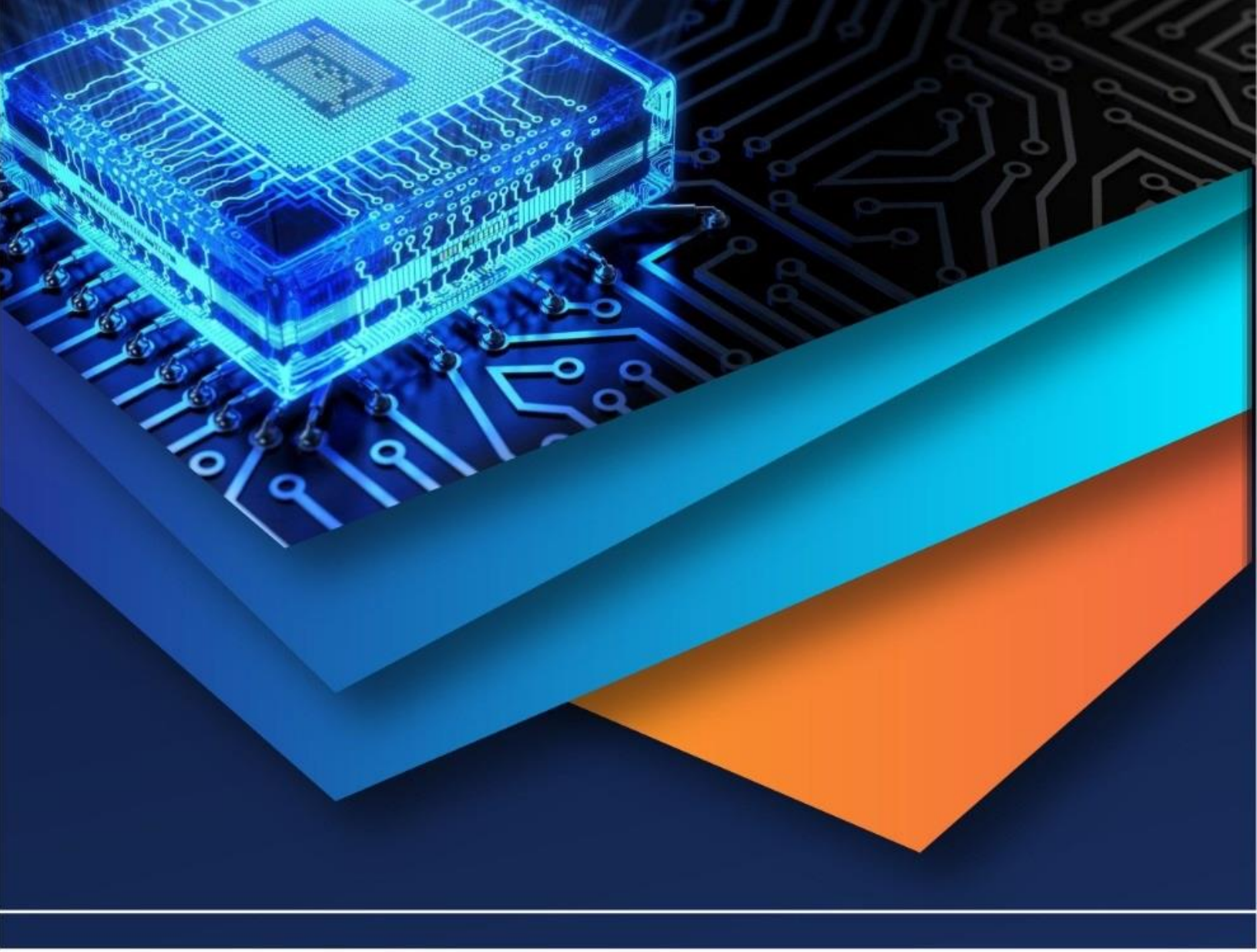
Gudadhe et al.[12] proposed a heart disease diagnosis model using multi-layer Perceptron and support vector machine (SVM) algorithms. This achieved accuracy of 80.41%. Humar et al.[13] designed diagnosis system by using a neural network with the integration of Fuzzy logic. The classification system achieved 87.4% accuracy. Palaniappan et al.[14] suggested an expert system for heart disease identification. The classifier model is made by using naïve bays (NB), Decision Tree (DT), and Artificial Neural Network (ANN). This model achieved the accuracy of 86.12%, 88.12% and 80.4% for naïve bays (NB), Artificial Neural Network (ANN) and Decision Tree (DT) respectively.

III.CONCLUSIONS

Heart is so dangerous that if left untreated, can lead to death. Or, can give irreversible effects on the health. That's why early detection or diagnosis is needed. But, traditional method requires time. Machine learning model for prediction of heart disease are faster and accurate than the traditional method at its early stages. Hybrid methods and ensemble methods improved prediction accuracy. DBSCAN-SMOTEEN-XGBoost together performed best among other techniques. This technique have the highest accuracy for Cleveland dataset of 98.40%.

REFERENCES

[1] World Health Organization. (2017). Cardiovascular Diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
[2] Li, Jian Ping, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare." IEEE Access 8 (2020): 107562-107582.
[3] Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System." IEEE Access 8 (2020): 133034-133050.
[4] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE Access 7 (2019): 81542-81554.
[5] Gupta, Ankur, Rahul Kumar, Harkirat Singh Arora, and Balasubramanian Raman. "MIFH: A machine intelligence framework for heart disease diagnosis." IEEE Access 8 (2019): 14659-14674.
[6] Javeed, Ashir, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor, and Redhwan Nour. "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection." IEEE Access 7 (2019): 180235-180243.
[7] Ali, Liaqat, Awais Niamat, Javed Ali Khan, Noorbakhsh Amiri Golilarz, Xiong Xingzhong, Adeeb Noor, Redhwan Nour, and Syed Ahmad Chan Bukhari. "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." IEEE Access 7 (2019): 54007-54014.
[8] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." Mobile Information Systems 2018 (2018).
[9] Shah, Syed Muhammad Saqlain, Safeera Batool, Imran Khan, Muhammad Usman Ashraf, Syed Hussnain Abbas, and Syed Adnan Hussain. "Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis." Physica A: Statistical Mechanics and its Applications 482 (2017): 796-807.
[10] Latha, C. Beulah Christalin, and S. Carolin Jeeva. "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques." Informatics in Medicine Unlocked 16 (2019): 100203.
[11] Nahato, Kindie Biredagn, Khanna Nehemiah Harichandran, and Kannan Arputharaj. "Knowledge mining from clinical datasets using rough sets and backpropagation neural network." Computational and mathematical methods in medicine 2015 (2015).
[12] Gudadhe, Mrudula, Kapil Wankhade, and Snehlata Dongre. "Decision support system for heart disease based on support vector machine and artificial neural network." In 2010 International Conference on Computer and Communication Technology (ICCCT), pp. 741-745. IEEE, 2010.
[13] Kahramanli, Humar, and Novruz Allahverdi. "Design of a hybrid system for the diabetes and heart diseases." Expert systems with applications 35, no. 1-2 (2008): 82-89
[14] Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." In 2008 IEEE/ACS international conference on computer systems and applications, pp. 108-115. IEEE, 2008.
[15] UCI Machine Learning Repository Heart Disease Data Set (1988). https://archive.ics.uci.edu/ml/datasets/Heart+Disease.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY