iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35626

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 VI Jun 2021- Available at www.ijraset.com

Review Paper on Prediction of Heart Disease using Machine Learning Algorithms

Aadar Pandita[1], Siddharth Vashisht[2], Aryan Tyagi[3], Prof. Sarita Yadav[4]
[1, 2, 3, 4]*Department of Information Technology, Bharati Vidyapeeth's College of Engineering, New Delhi*

Abstract: Heart disease has been one of the ruling causes for death for quite some time now. About 31% of all deaths every year in the world take place as a result of cardiovascular diseases [1].
A majority of the patients remain uninformed of their symptoms until quite late while others find it difficult to minimise the effects of risk factors that cause heart diseases.
Machine Learning Algorithms have been quite efficacious in producing results with a high level of correctness thereby preventing the onset of heart diseases in many patients and reducing the impact in the ones that are already affected by such diseases. It has helped medical researchers and doctors all over the world in recognising patterns in the patients resulting in early detections of heart diseases.
Keywords: Cardiovascular Diseases (CVDs); Support Vector Machine (SVM); K- Nearest Neighbour (KNN); Naive Bayes (NB); Random Forest (RF); Logistic Regression (LR); Machine Learning (ML); Prediction Model

I. INTRODUCTION

Machine Learning being a subset of Artificial Intelligence has been one of the most vital tools in recent history. It has helped further the progress of the healthcare industry manifolds. With such a tool, medical professionals and researchers have been able to diagnose and detect diseases with much accurate precision. It has contributed to saving many lives.
Majority of the heart related diseases can be averted if people focus on their physical activity, have a balanced and nutritious diet and avoid consumption of products such as tobacco and alcohol that have a damaging effect on the heart and their overall health. Therefore, it becomes paramount that these diseases are detected as early, so that its effects can be managed with medical advice and medicines. In this paper, we are going to review various advancements and recent works that have been done using Machine learning in the prediction of heart diseases. Heart diseases are a result of a multitude of aspects that can influence the cardiovascular health of an individual such as age, blood sugar, blood pressure, cholesterol etc.

II. LITERATURE SURVEY

Apurv Garg et al. [2] implemented KNN and Random Forest machine learning algorithms in order to predict heart diseases. After obtaining and analysing the data, its balancing was checked and correlation was found between various attributes and their effect on the target value. The dataset obtained was the UCI dataset available at Kaggle. It was divided in 80-20 ratio for training and testing respectively. It was found that Chest Pain and Maximum heart rate achieved had a positive correlation with the target attribute. This model provided an accuracy of 86.885% using KNN and 81.967% accuracy using Random Forest.
Rishabh Magar et al. [3] proposed a web application based predictive model trained on the UCI dataset with a 75-25 training and testing division of the dataset. Logistic Regression based predictive models were found to be the most accurate with 82.89% accuracy, followed by SVM at 81.57% and Naive Bayes and Decision Tree at 80.43% each. The web application can be used by the end user as a preliminary test for checking their heart condition and seeking medical advice if needed.
Apurb Rajdhan et al. [4] proposed a system where four classification algorithms such as Random Forest, Decision Tree, Logistic Regression and Naive Bayes are used to predict the patient's condition. Data is split into 80% training data and 20% testing data. A confusion matrix depicting true and false positives as well as true and false negatives was created. Maximum accuracy obtained was 90.16% using Random Forest classification.
Devansh Shah et al. [5] proposed a system of models using supervised learning methods through the WEKA tool. Four individual classification techniques including NB, KNN, RF, DT were used to predict the chances of having a heart disease. The dataset was initially cleaned, transformed by smoothening, normalisation, and aggregation, integrated and reduced. The maximum accuracy obtained was through KNN method.

Harshit Jindal et al. [6] implemented a system that uses three different classification algorithms, KNN, RF, LR and results in an accuracy of 87.5%. In this EHDPS i.e, effective heart disease prediction system, Logistic Regression and KNN outperform RF with KNN providing an accuracy of 88.52% which is highest amongst the three techniques used.

Aadar Pandita et al. [7] proposed a predictive model that implements 5 machine learning algorithms and uses the technique with the highest accuracy to build a web application that takes in patient's medical details and predicts if they have a heart disease or not. The web application is built using HTML/CSS and Flask based framework. The maximum accuracy obtained was obtained using KNN, i.e, 89.06% while Logistic Regression contributed with least accuracy of 84.38%.

N. Saranya et al. [8] proposed a time and money efficient model of predicting heart disease using a web application. The model works on two different methods : Random Forest and KNN. The dataset has been taken from one of Coimbatore's hospitals which produces an accuracy of 100% using Random Forest and 91.36% using KNN after cleaning and pre-processing of the dataset. An ensemble model with and without Logistic Regression is also used to predict the chances with an accuracy of 98.77% and 95.06% respectively.

Aravind Akella et al. [9] applied six predictive models on the UCI dataset and achieved a maximum accuracy of 93.03% with Neural Networks with a recall of 93.8 indicating low chances of false negatives and therefore extremely precise result, while the other five had an accuracy of almost 80% and more.

Ravindhar NV et al. [10] implemented five algorithms: Logistic Regression, Naive Bayes, Fuzzy KNN, K-Means Clustering and back propagation Neural-Network. A 10-fold cross validation method is used in the experimental analysis of heart conditions. The maximum accuracy was gathered using back propagation Neural Network with 98.2% accuracy and 87.64% recall and 89.65% precision.

A comparative study of various papers on heart disease prediction using ML models

Year	Author	Paper Name	Algorithms Used	Accuracy Obtained
2020	Rishabh Magar et al. [3]	Heart disease prediction using machine learning	Logistic Regression SVM Naive Bayes Decision Tree	82.89% 81.57% 80.43% 80.43%
2020	Apurb Rajdhan et al. [4]	Heart disease prediction using machine learning	Logistic Regression Decision Tree Random Forest Naive Bayes	85.25% 81.97% 90.16% 85.25%
2020	Devansh Shah et al. [5]	Heart disease prediction using machine learning techniques	Naive Bayes KNN Random Forest Decision Tree	88.157% 90.789% 86.84% 80.263%
2021	Harshit Jindal et al. [6]	Heart Disease prediction using machine learning algorithms	KNN Logistic Regression KNN & LR based model	88.52% 88.5% 87.5%

2021	Aadar Pandita et al. [7]	Prediction of Heart Disease using Machine Learning Algorithm	Logistic Regression KNN SVM Naive Bayes Random Forest	84.38% 89.06% 87.50% 85.94% 87.50%
2020	N. Saranya et al. [8]	Heart Disease prediction using Machine Learning	Random Forest KNN Logistic Regression Ensemble model with Logistic Regression Ensemble model without Logistic Regression	100% 91.36% 87.65% 95.06% 98.77%
2021	Aravind Akella et al. [9]	Machine learning algorithms for predicting coronary artery disease : efforts toward an open source solution	Generalized linear model Decision Tree Random Forest SVM Neural Network KNN	87.64% 79.78% 87.64% 86.52% 93.03% 84.27%
2019	Ravindhar NV et al. [10]	Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning	Logistic Regression Naive Bayes Fuzzy KNN K-Means Clustering BP-Neural Network	81.86% 61.46% 87.33% 43.24% 98.20%

III. CONCLUSION

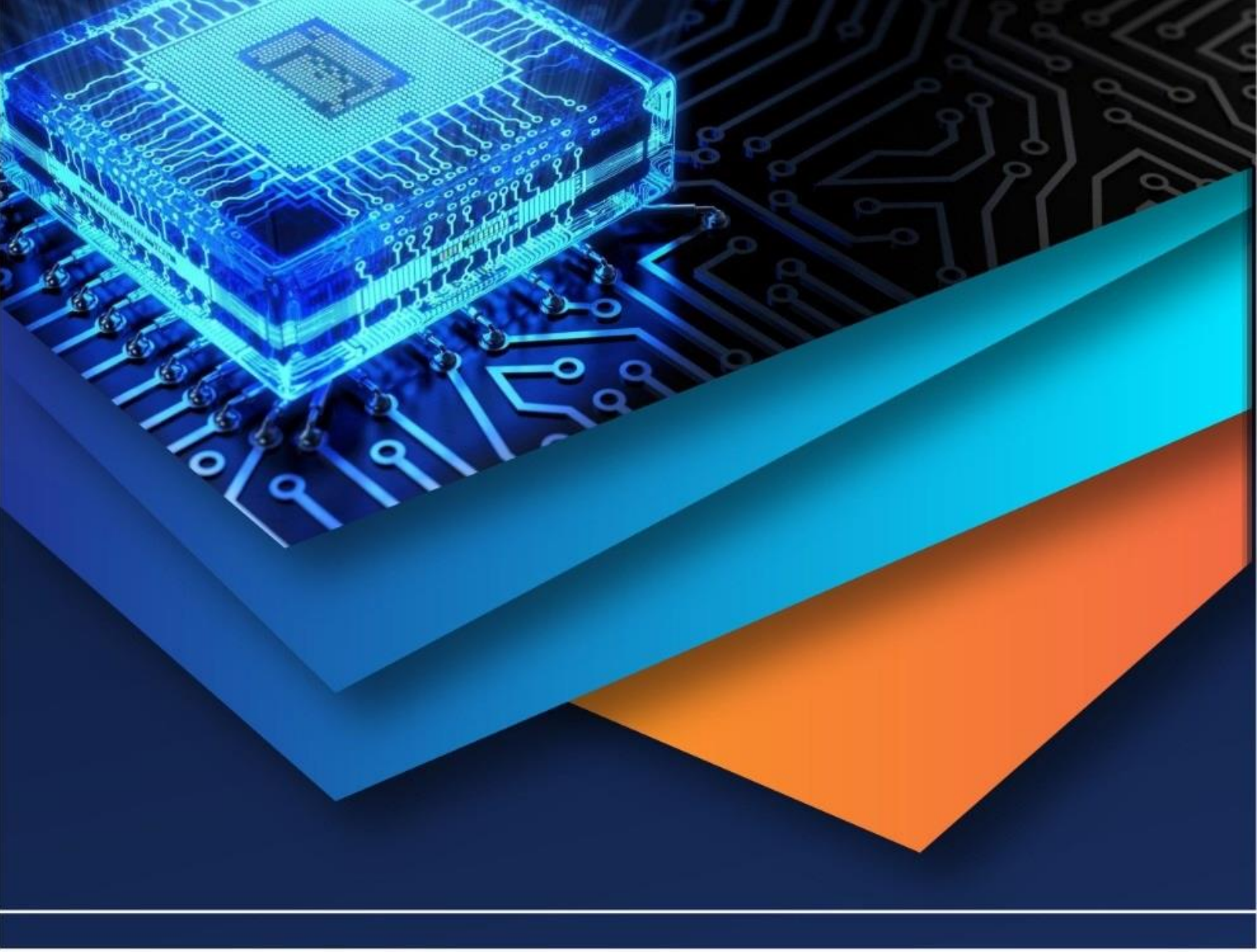
Majority of researchers have used the Cleveland Heart Disease Dataset available from the UCI repository containing 76 attributes and 303 instances, of which only 14 attributes are used due to missing values [11]. There are huge benefits to having feature selection methods so as to minimise the number of attributes that one has to use in order to build an accurate model by checking the correlation between various attributes and their impact on the accuracy of the models. It can be seen from various research papers in the field that KNN and Neural Network works quite accurately in most cases for the prediction of heart diseases.

IV. FUTURE WORK

Future researchers should work towards improving the existing accuracies. They can create their own dataset using the existing datasets available in order to increase the sample size and allow predictive models to train on a larger dataset thereby increasing the chances of obtaining improved accuracy. Special focus should be put towards removing false positives and false negatives from the existing models. The predictive models should be accessible to the people in the form of a web or a mobile application so that people can try to be aware of their heart condition and consult a medical professional if their results predict a related disease. Although users of such applications should keep in mind that these applications are not a substitute for a doctor, rather just a model trained on a certain dataset which could have errors.

REFERENCES

[1] https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] Garg, Apurv & Sharma, Bhartendu & Khan, Rizwan. (2021). Heart disease prediction using machine learning techniques. IOP Conference Series: Materials Science and Engineering. 1022. 012046. 10.1088/1757-899X/1022/1/012046.
[3] "HEART DISEASE PREDICTION USING MACHINE LEARNING", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.2081-2085, June-2020, Available :http://www.jetir.org/papers/JETIR2006301.pdf
[4] Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020)
[5] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020). https://doi.org/10.1007/s42979-020-00365-
[6] Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072
[7] Aadar Pandita, Siddharth Vashisht, Aryan Tyagi, Prof. Sarita Yadav."Prediction of Heart Disease using Machine Learning Algorithms", Volume 9, Issue V, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 2422-2429, ISSN : 2321-9653, www.ijraset.com
[8] N. Saranya, P. Kaviyarasu, A. Keerthana, C. Oveya. Heart Disease Prediction using Machine Learning International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-1, May 2020, Page No: 700-70
[9] Akella, Aravind and Akella, Sudheer. Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. Future Science OA Volume 7, Number 6, Pages FSO698, 2021, https://doi.org/10.2144/fsoa-2020-0206
[10] Ravindhar NV, Anand, Hariharan Shanmugasundaram, Ragavendran, Godfrey Winster. Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning. Volume-8 Issue-11, September 2019, ISSN: 2278-3075 (Online). Page No: 1417-1421. DOI: 10.35940/ijitee.J9765.0981119
[11] 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