ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE
Your ub-MATE Research Paper !!!

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 5, Issue 1, January 2019

Prediction of Heart Disease Using Machine Learning

SURESH B
Senior Grade Lecturer in CSE
Government Polytechnic Koppal
suresh.arb@gmail.com

P. R ASHALATHA
Senior Grade Lecturer in CSE
Government Polytechnic K.R.Pete
ashl12.pr@gmail.com

CHANNAPPA A.
Senior Grade Lecturer in CSE
Government Polytechnic Kudligi
channappajeevitha@gmail.com

Abstract: Heart disease, often known as cardiovascular disease, is the leading cause of death globally over the past few decades. It includes a variety of disorders that have an impact on the heart. Numerous risk factors for heart disease are linked to the requirement for timely access to accurate, trustworthy, and practical methods for early diagnosis and disease management. Data mining is a popular method for processing vast amounts of data in the healthcare industry. In order to forecast cardiac disease, researchers analyze vast amounts of complex medical data using a variety of data mining and machine learning techniques.

In this study, numerous heart disease-related characteristics are presented, along with a model built using supervised learning techniques such Naive Bayes, logistic regression, Neural network and random forest algorithm.It makes use of the current dataset from the UCI heart disease patient repository's Cleveland, Hungary, Switzerland database. There are 920 instances and we have taken crucial 14 attributes to proving the effectiveness of various algorithms—are taken into consideration for testing. The purpose of this study work is to predict the likelihood that patients would develop heart disease. The findings show that Neural networks yields the highest accuracy score.

Keywords: Heart disease prediction · Logistic regression · Naïve Bayes · Neural Network · Random forest · Machine learning

Introduction

The main cause of death globally. According to a World Health Organization estimate, about 17.9 million people worldwide die each year from cardiovascular disease, with coronary artery disease and cerebral stroke accounting for 80% of these deaths[1]. Low and middle income nations frequently have a high death rate. [2] discussed that Biomedical and anatomical data are made simple to acquire because of progress accomplished in computerizing picture division. More research and work on it has improved more viability to the extent the subject is concerned. A few tech- niques are utilized for therapeutic picture division, for example, Clustering strategies, Thresholding technique, Classifier, Region Growing, Deformable Model, Markov Random Model and so forth. This work has for the most part centered consideration around Clustering techniques, particularly k-implies what's more, fluffy c-implies grouping calculations. These calculations were joined together to concoct another technique called fluffy k-c-implies bunching calculation, which has a superior outco- me as far as time usage. The calculations have been actualized and tried with Magnetic Resonance Image (MRI) pictures of Human cerebrum. The proposed strategy has expanded effectiveness and lessened emphasis when contrasted with different techniques. The nature of picture is assessed by figu- ring the proficiency as far as number of rounds and the time which the picture takes to make one emphasis. Results have been dissected and recorded. Some different strategies were surveyed and favorable circumstances and hindrances have been expressed as special to each. Terms which need to do with picture division have been characterized nearby with other grouping strategies. Heart disease is caused by a variety of risk factors, including genetic predisposition, personal and professional habits, and lifestyle choices. Cardiac disease is predisposed to by a number of behavioural risk factors, including smoking, excessive alcohol and caffeine use, stress, and physical inactivity, in addition to physiological risk factors like obesity, hypertension, high blood cholesterol, and pre-existing heart diseases. In order to take action to save death, an early, accurate and effective medical diagnosis of heart disease is essential.Researchers have conducted a plethora of studies in an effort to lower the rates of morbidity and mortality from heart disease, which are on the rise throughout the world. [3] proposed a principle in which the division is the urgent stage in iris acknowledgment. We have utilized the worldwide limit an incentive for division. In the above calculation we have not considered the eyelid and eyelashes relics, which corrupt the execution of iris acknowledgment framework. The framework gives sufficient execution likewise the outcomes are attractive. Assist advancement of this technique is under way and the outcomes will be accounted for sooner rather than later. Based on the reasonable peculiarity of the iris designs we can anticipate that iris acknowledgment framework will turn into the main innovation in personality verification.In this paper, iris acknowledgment calculation is depicted. As innovation advances and data and scholarly properties are needed by numerous unapproved work force.

Therefore numerous associations have being scanning routes for more secure confirmation strategies for the client get to. The framework steps are catching iris designs; deciding the area of iris limits; changing over the iris limit to the binarized picture; The framework has been actualized and tried utilizing dataset of number of tests of iris information with various complexity quality.

One of the areas of artificial intelligence that is advancing the fastest is machine learning. These algorithms are capable of analyzing vast amounts of data from many different sectors, the medical field being one of them.By lowering the errors in projected and actual outcomes, it is a replacement for conventional prediction modelling approach employing a computer to analyse complicated and non-linear interactions among many elements[4].

Huge datasets are analysed through data mining in order to uncover hidden, vital decision-making information from a repository of the past for future research. The amount of patient data in the medical industry is enormous. Various machine learning algorithms must mine these data. These data are analysed by healthcare specialists to help them make good diagnostic decisions. Analyzing medical data using categorization algorithms offers clinical assistance. In order to forecast cardiac disease in patients, it evaluates the categorization algorithms [5].

The technique of removing important data and information from sizable databases is known as data mining. Heart disease is predicted using a variety of data mining approaches, including regression, clustering, association rules, and classification techniques including Naive Bayes, decision trees, random forests, and K-nearest neighbours.It uses a comparative examination of the various classification methods[6].

We used data from the UCI repository for this study. Using classification methods for heart disease prediction, the classification model is created. This study compares the current methods and discusses the algorithms used for heart disease prediction.The study also discusses opportunities for advancement and additional research.

Methodology

Machine learning is essential in healthcare organizations for automating systems and enhancing the working environment. DM helps to improve service quality while cutting costs, for example, the heart disease prediction will enable the clinician to more precisely define the illness[7]. In today's healthcare facilities, a sizable volume of data is handled electronically, making traditional analysis unfeasible[8]. Additionally, software can be used to analyse enormous amounts of data in databases or other information repositories in order to save lives[9].

In order to help doctors and patients in the medical industry, this study attempts to forecast the likelihood of having heart disease as a likely cause of computerised heart disease prediction [5]. In this research study, we address the application of several machine learning algorithms to the dataset in order to achieve the goal. We also discuss dataset analysis. This paper also illustrates certain characteristics are more important than others in predicting higher precision. This could save patient money on various trials because not all of their characteristics will necessarily have a significant impact on the outcome [5].

Data set used

I used data from the UCI Machine Learning repository for this project. It includes an actual dataset of 920 examples of data with 14 different attributes (13 predictors; 1 class), such as blood pressure, the nature of chest discomfort, the outcome of an ECG, etcas described in Table 1. In order to construct a model with the highest level of accuracy feasible, we applied four algorithms in this study to determine the causes of heart disease.

Table 1 Attributes and details of dataset of heart disease

Sr. no	Attributes	Description
1	age	Age Patients age, in years
2	sex	0 = female; 1 = male
3	cp	Chest pain ,types of chest pain (1—typical angina; 2—atypical angina; 3—non-angina pain; 4—asymptomatic)
4	trestbps	Rest blood pressureResting systolic blood pressure (in mm Hg on admission to the hospital)
5	chol	Serum cholesterol Chol Serum cholesterol in mg/dl
6	fbs	Fasting blood sugar Fbs Fasting blood sugar > 120 mg/dl (0—false; 1—true)
7	restecg	Rest electrocardiograph 0—normal; 1—having ST-T wave abnormality; 2—left ventricular hypertrophy
8	thalch	Maximum heart rate achieved
9	exang	Exercise-induced angina (0—no; 1—yes)
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	slope of the peak exercise ST segment (1—upsloping; 2—flat; 3—down sloping)
12	ca	No. of major vessels (0–3) colored by fluoroscopy)
13	thal	Defect types; 3—normal; 6—fixed defect; 7—reversible defect
14	num	diagnosis of heart disease status (0—nil risk; 1—low risk; 2—potential

		risk; 3— high risk; 4—very high risk)

Machine learning classification algorithms namely Neural Network, Logistic Regression, Naive Bayes, and Random Forest were applied on the dataset.

The dataset is divided into sections for testing and training. Due of its improved accuracy, the K-Fold-Cross-Validation technique is frequently employed. The data is divided into M folds using the K-Fold-Cross-Validation technique, where M folds are used for learning and the M^{th} fold is used for testing the data. The data operates on various percentage divisions. By setting the sub-samples number to 10, this technique was utilised to confirm the analysis's standard. By contrasting the outcomes of the procedures using different parameters, the system is assessed.The random forest is implemented by taking about 10 trees and neural network model is built with 100 neuron hidden layers, ReLu activation function.

Different settings and parameters were used to evaluate the credit risk assessment approaches. Common metrics including F1 score, specificity, accuracy, sensitivity, error rate, and precision were used to compare these approaches.The results of the various methods used to assess the credit risk are listed in Table II and the accuracy obtained the different models is shown in figure1.

Table II: The performance of the models

Model	**AUC**	**F1**	**Precision**	**Recall**
Random Forest	0.88	0.78	0.74	0.82
Neural Network	0.89	0.80	0.77	0.82
Naive Bayes	0.88	0.78	0.83	0.73
Logistic Regression	0.88	0.80	0.74	0.87

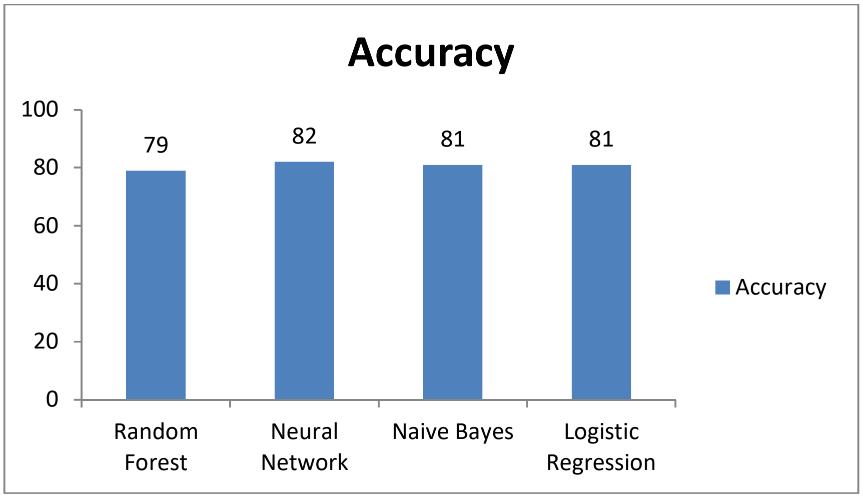

Fig 1: Accuracy of the different models

Results and Analysis

The purpose of this study is to forecast a patient's risk of developing heart disease. On the UCI repository, this study used supervised machine learning classification methods such as Naive Bayes, decision trees, random forests, and K-nearest neighbors. Several trials employing various classifier algorithms were carried out. The highest accuracy of 82% is obtained for neural network model.

Conclusion

The overarching goal is to define several machine learning methods that can be effectively used to forecast cardiac disease. Our objective is efficient and accurate prediction using fewer features and tests. We just take into account 14 key characteristics in our study. Logistic regression, Naive Bayes, Random forest and neural network were the classification algorithms.After employing four algorithms, it was discovered that Neural Network had the highest accuracy.

Additional data mining techniques, such as time series, clustering and association rules, support vector machines, and genetic algorithms, can be used into this research to further its scope. The limitations of this study need the adoption of more intricate and coupled models in order to increase the precision of heart disease early prediction.

Reference

[1] M. D. Seckeler and T. R. Hoke, "The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease," *Clinical epidemiology,* vol. 3, p. 67, 2011.

[2] Christo Ananth, S.Aaron James, Anand Nayyar, S.Benjamin Arul, M.Jenish Dev, “Enhancing Segmentation Approaches from GC-OAAM and MTANN to FUZZY K-C-MEANS”, Investigacion Clinica, Volume 59, No. 1, 2018,(129-138).

[3] Christo Ananth,"Iris Recognition Using Active Contours",International Journal of Advanced Research in Innovative Discoveries in Engineering and Applications[IJARIDEA],Volume 2,Issue 1,February 2017,pp:27-32.

[4] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PloS one,* vol. 12, p. e0174944, 2017.

[5] V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," *International Journal of Engineering & Technology,* vol. 7, pp. 684-687, 2018.

[6] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science,* vol. 1, pp. 1-6, 2018.

[7] A. Taneja, "Heart disease prediction system using data mining techniques," *Oriental Journal of Computer science and technology,* vol. 6, pp. 457-466, 2013.

[8] V. Chaurasia and S. Pal, "Early prediction of heart diseases using data mining techniques," *Caribbean Journal of Science and Technology,* vol. 1, pp. 208-217, 2013.

[9] M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of heart failure by using unstructured data of cardiac patients," in *2016 45th International Conference on Parallel Processing Workshops (ICPPW)*, 2016, pp. 426-431.