International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2390439

Heart Risk Prediction using Machine Learning : A Literature Review

Om Deshmukh[1], Fardeen Kachawa[1], Sujal Bhatt[1], Kaif Siddique[1], Bhavesh Choudhary[1], Neelam Phadnis[2]

[1]Computer Engineering Department, Student, Shree L.R. Tiwari College of Engineering, Mira Road, Thane, Maharashtra, India

[2]Computer Engineering Department, Assistant Professor, Shree L.R. Tiwari College of Engineering, Mira Road, Thane, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 Aug 2023

Published: 27 Aug 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

409-413

ABSTRACT

Heart diseases are a leading cause of death among people compared to other diseases. The severity of these diseases has risen significantly in the past few years which has led to the rise of many researchers to present their work in the field of heart risk detection. Machine learning plays an important role in this with the most common machine learning algorithms used for this purpose being Logistic Regression, Naive Bayes, SVM, etc. All these algorithms fall under the classification algorithm category. Data mining plays an important role for feature selection from the dataset. The machine learning algorithms reviewed make use of the same UCI Cleveland dataset.

Keywords : Machine Learning, Logistic Regression, Naive Bayes, SVM, Classification Algorithm, Data Mining

I. INTRODUCTION

According to the World Health Organisation (WHO) report, stroke and heart attack constitute around 80% of global Cardiovascular related death [1]. Heart Diseases or Cardiovascular Diseases (CVDs), also referred to as silent killers, are the leading cause of disease burden and mortality across the globe. According to the World Health Organisation estimation by 2030, very nearly 23.6 million individuals will pass away because of Heart diseases. So to minimise the danger, an expectation of coronary disease ought to be finished. Machine Learning plays a very important role to detect hidden discrete patterns and thereby analyse the provided data. After analysis of data, Machine Learning techniques help in heart disease prediction and early diagnosis.

II. BACKGROUND

Treatment for heart diseases is crucial as there has been an increasing number of cases that suggest the rising trend of these issues among people particularly in a developing nation such as India. In 2000, there has been an estimate 29.8 million people suffering from coronary heart disease (CHD) which equates to 3% at that time [2]. In 2003, that number increased significantly as the CHD affected population number increased to 5% [3]. According to WHO country wise statistics, 53% of fatalities in India are a result of Non communicable diseases of which 24% are due to CVDs.

The following are the most prevalent heat-related illnesses in India right now [4]:

A. Ischemic heart disease

The term "ischemic heart disease," also known as "coronary heart disease" or "coronary artery disease," refers to cardiac conditions brought on by clogged coronary arteries which are responsible to deliver blood to the heart muscle. [5]

B. Cerebrovascular Disease

An abrupt deterioration of the cerebral perfusion or vasculature connected directly to the cardiovascular network is referred to as a stroke or cerebrovascular accident (CVA). Strokes are ischemic in about 85% of cases, and hemorrhagic in the remaining 15%[7].

C. Rheumatic Heart Disease

Rheumatic fever, an inflammatory disease that can develop when strep throat or scarlet fever are not appropriately treated, can lead to rheumatic heart disease, a systemic immunological syndrome. A serious form of acquired heart disease that affects both children and adults globally is rheumatic heart disease [6].

D. Hypertensive heart disease

Chronically elevated blood pressure puts more strain on the heart, causing structural and functional changes that are referred to as hypertensive heart disease. These alterations affect the left ventricle, left atrium, and coronary arteries [7].

E. Cardiomyopathy

Anatomically and pathologically, cardiomyopathy refers to heart muscle or electrical malfunction. Cardiomyopathies are a diverse set of illnesses that frequently result in progressive heart failure and have high morbidity and mortality rates [1].

F. Atrial fibrillation

The most common heart arrhythmia, atrial fibrillation (AF), affects 1% to 2% of the general population. The absence of a P-wave and erratic QRS complexes on an EKG can be used to detect this condition, which is characterised by fast and chaotic atrial activation that results in compromised atrial function [8].

III. NEED OF DATA MINING

For the current problem, data mining is of utmost importance as it is responsible for the selection of test and training data for the machine learning model which are necessary for the most accurate prediction of presence of any heart risks [9].

Data cleaning is the main part of pre-processing which is classified into two main categories:

i) Removal of duplicate data:

A lot of instances that consists of data values in any given dataset to be repeated. In such cases, we must eliminate them as they can lead to discrepancies in our training model. As suggested in [13] [10], we form a number of rules for the data values to ensure that only the necessary data can be picked for the model, the data values that conform to these rules are used further, every other data value will be treated as error and eliminated from the dataset.

ii) Error Repairing

Error repairing is of various kinds with the prominent method that is still used to this day is manually fixing the incorrect data value. Another method employed to perform this task is to automate this entire process by making changes into the databases[13].

These techniques are necessary to make sure that the dataset excludes all the outliers and errors from the dataset allowing us to work with it more effectively.

In case of Heart Risk predictions, we would be left with the following data :

1) age in years
2) sex(1 = male, 0 = female)
3) chest pain type
4) serum
5) cholesterol

6) fasting blood sugar
7) resting electrocardiographic result
8) maximum heart rate achieved
9) exercise included angina(exhang)
10) old peak
11) slope
12) number of major vessel
13) thal
14) resting blood pressure

These parameters are taken from the Kaggle UCI heart disease dataset as well as Cleveland UCI heart disease dataset[14].
Majority of the proposed systems that we reviewed utilised the above given dataset with the same parameters.

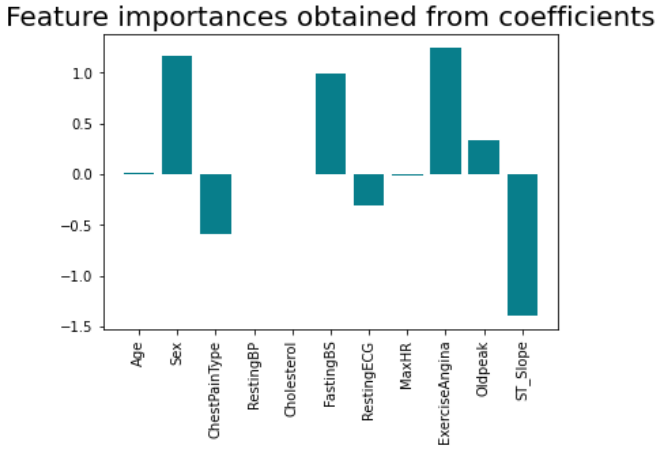

Fig 1. Feature Importance

figure 1 represents the importance of each parameter in the dataset that decides whether a certain parameter can result in presence of heart risk.

For the current topic, we would be reviewing the previous research on this topic to outline which method is the most suitable for predicting heart risks.

Classification is the process of learning a function that can classify data objects to a subset of a given class set. Finding a decent generic that can accurately forecast the class of as- yet-unknown data items is one of the classification's first types of tasks. The next step is to locate a condensed and simple to comprehend class model for each class [11].

IV. LITERATURE REVIEW

Due to the rise of heart diseases in the world, many researchers stepped forward to find a way to predict these risks using existing algorithms at times with certain extensions.
Pooja Anbuselvan [12]studied and implemented various machine learning algorithms on the UCI heart disease dataset. The logistic regression algorithm fetched an accuracy of 75.41% after data cleaning which included removing incomplete records altogether.
Vembandasamy K. ,Sasipriya R. and Deepa E. [13] proposed and implemented the naive bayes algorithm on the database of one of the leading diabetes research institutes in chennai which acquired them with an accuracy of 74% .

Mai Shouman, Tim Turner and Rob Stocker [14] have [4] R. Gupta, “Burden of coronary heart disease in implemented the decision tree algorithm, namely the j48 and India,” Portal Regional da BVS, vol. 57, no. 6, bagging algorithm type of the decision tree algorithm on the 2005.
Cleveland Clinic Foundation Heart disease data set which acquired them an accuracy of 78.9% on j48 and 81.41% on [5] C. Dass and A. Kanmanthareddy , “Rheumatic Heart Disease,” StatPearls Publishing. bagging algorithm respectively.
M.A.Jabbar , B.L.Deekshatulu and Priti Chandra [15]has implemented a hybrid machine learning model of random forest with chi squared method and genetic algorithm that they applied on the heart disease dataset of the patients in corporate hospitals which fetched them an accuracy of 83.70%.
Anuradha.P and Dr.Vasantha Kalyani David implemented XGgradient boosting algorithm on multiple cleveland dataset after feature selection based

on information gain which fetched them an accuracy of 88.52%.

V. RESULTS AND DISCUSSION

ALGORITHMS	ACCURACY
Logistic Regression	75.41%
Naive Bayes	74%
Decision Tree	81.41%
Random Forest	83.70%
XGgradient Boosting	88.52%

The table shows us the accuracy scores of each algorithm as per the authors.

XGgradient Boosting algorithm has the highest accuracy of all the compared algorithms.

VI. REFERENCES

[1]. www.who.int/cardiovascular_diseases/en/ [Accessed 20 May 2023].

[2]. S. Chauhan and . B. T. Aeri, "The rising incidence of cardiovascular diseases in India:," ResearchGate, vol. 4, no. 4, May 2015.

[3]. R. Gupta, J. P, M. V, K. S. Reddy and S. Yusuf, "Epidemiology and causation of coronary heart disease and stroke in India," BMJ Journals, vol. 94, no. 1, pp. 16-26, 2007.

[4]. G. Tackling and M. B. Borhade, "Hypertensive Heart Disease," National Library of Medicine, 27 June 2022. [Online]. Available:https://www.ncbi.nlm.nih.gov/books/NBK539800/. [Accessed May 2023].

[5]. R. Wexler, T. Elton, A. Pleister and D. Feldman, "Cardiomyopathy: An Overview," National Library of Medicine, 9 December 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2999879/. [Accessed May 2023].

[6]. J. Pellman and F. Sheikh, "Atrial Fibrillation: Mechanisms, Therapeutics, and Future Directions," National Library of Medicine, 17

[7]. January 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5240842/. [Accessed May 2023].

[8]. N. Jothi, N. A. Rashid and W. Husain, "Data Mining in Healthcare – A Review," ScienceDirect, vol. 72, pp. 306-313, 2015.

[9]. X. Chu, I. F. Ilyas, S. Krishnan and J. Wang, "Data Cleaning: Overview and Emerging Challenges," Association for Computing machinery, pp. 2201-2206, 26 June 2016.

[10]. C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," in IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputtur, 2017.

[11]. P. Anbuselvan, "Heart Disease Prediction using Machine Learning," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 11, pp. 515-518, 2020.

[12]. K. Vembandasamy, R. Sasipriya and E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm," - International Journal of Innovative Science, Engineering & Technology, vol. 2, no. 9, pp. 441-444, 2015.

[13]. M. Shouman, T. Turner and R. Stocker, "Using Decision Tree for Diagnosing Heart Disease Patients," in Australasian Data Mining Conference, Ballarat, 2011.

[14]. M. A. Jabbar, B. L. Deekshatulu and P. Chandra, "Intelligent heart disease prediction system using," Journal of Network and Innovative Computing, vol. 4, pp. 175-184, 2016.

[15]. A. P and V. K. David, "Feature Selection and Prediction of Heart diseases using Gradient Boosting Algorithms," in International Conference on Artificial Intelligence and Smart Systems, Coimbatore, 2021.

[16]. D. Prabhakaran, P. Jeemon and A. Roy, “Cardiovascular Diseases in India,” AHA Journals, vol. 133, no. 16, 2016.

[17]. A. S. Khaku, P. Tadi and A. A. Gunn, “Cerebrovascular Disease,” National library of Mediciine, 2022.

[18]. M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn and M. A. Moni, “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,” ScienceDirect, vol. 136, 2021.

Cite this article as :

Om Deshmukh, Fardeen Kachawa, Sujal Bhatt, Kaif Siddique, Bhavesh Choudhary, Neelam Phadnis, "Heart Risk Prediction using Machine Learning : A Literature Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.409-413, July-August-2023.
Available at doi :
https://doi.org/10.32628/CSEIT2390439
Journal URL : https://ijsrcseit.com/CSEIT2390439