International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Comparative Analysis of AI Techniques in the Prediction of Heart Disease

Irtiqa Dhar

Department of Computer Science Engineering, Mewar University, Rajasthan, India
dharirtiqa14 [at]gmail.com

Abstract: ***Heart disease is very common problem today, mostly deaths occur due to cardiac failure. Problems in Electrocardiogram, Lipid profile, Blood pressure, Pulse rate, these are the major signs of cardiovascular related problems. Machine learning an emerging field which plays an important role in the prediction of cardiovascular problems. There are number of AI techniques present, but in this context an analysis of few techniques has been done. The techniques used are Decision Tree, Neural Network, K Mean clustering.***

Keywords: Artificial Intelligent Techniques, supervised learning, classification, cluster, Triglyceride, Low-density lipoprotein

1. Introduction

Cardiovascular disease is the main reason for death today. Cardiovascular disease is a disease of heart and blood vessels. There are many factors like high pulse rate, sudden fall in body temperature, High LDL, High cholesterol, High Triglycerides, Smoking, High BP. Arteries narrow down due to accumulation of plaque and this type of heart problem is called Atherosclerosis. This directly attack on Left main coronary artery and the right coronary artery that results in a cardiac failure and in some cases a major heart stroke. Cardiac diseases are of different types via Congenital Heart disease, it is present in heart since birth. Arrhythmia- it is abnormality related to heartbeat. Bradycardia-it is another situation when heart beats very fast and Bradycardia -when heart beats very slowly. As mentioning above there are no. of reasons which are responsible for improper functioning of heart some have greater influence than others like 13% of the heart disease occur due to high BP. 6% of people die due to intake of tobacco, 5% of people die due to obesity [1]. In Asian countries, one fifth of lives are lost due to cancer, cardiac problems and diabetes [20]. Australia Bureau has recorded 33.7% of deaths occur due to cardio vascular disease [21]. Usually decisions made by medical professionals are based on experience or some intuitions by analysing the no. of reports which may lead to the unnecessary bias [9].

Artificial intelligence (AI) plays an important role in the medical field. Different AI techniques show different accuracy levels depending on the attributes, testing and training methods used to get the result [4]. AI has applications in other emerging fields like Robotics, Guiding population -level disease prevention, Remote sensing, Natural language processing.

This paper presents a comparison of classification techniques like Decision tree, Artificial Neural Networks, K mean clustering and the analysis has been done on the primary dataset that was collected from the Government Medical College Srinagar and Medical diagnostic centre Srinagar. Most interesting part is only two tests were analysed to perform the experiment and the rest of the attributes don't require any specific test. The tests that were considered are Lipid profile and ECG report.

Section II of the paper consists previous work that is the literature review. Section III presents methodology. Section IV highlights techniques which are implemented and the Section V consists experimental result followed by the future work and the conclusion.

2. Literature Review

- Anjinky Kunjir Harshalsawant, Nuzhatfsheikh [22] used Data mining and Visualisation for prediction of multiple heart disease in health care.
- Mai Shouman, Tim Turner, Rob Stocker [4] used K Nearest Neighbor on the Cleveland data and investigation shows that k-means could not perform but the accuracy can be increased by applying voting method in the prediction of heart disease.
- Salma banu N.K, Suma Swamy [6] used both data mining and big data analysis in the prediction of cardiovascular disease and technique depends on the type and number of attributes used
- N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra [16] presented CANFIS system that is Coactive Neuro Fuzzy Inference system, combination of neural networks, fuzzy logic and genetic algorithm in the diagnosis of cardiovascular diseases
- A. Dewan and M. Sharma implemented back propagation algorithm in the prediction of cardiovascular disease and performed better in comparison to other previous technique used.
- Sultana, Afrin Haider and Mohammad Shorif Uddin [5] applied Kstar, MLP, J48, BayesNet, SMO and used ROC curve for efficiency comparison, conducted on weka, Bayes Net and SMO performed better than other techniques used.

It is proved that all the Artificial Intelligence techniques which are mentioned above perform better in the prediction of any cardiovascular disease. Different techniques analysis differently and have different accuracy. In the above researches most of the analysis has been conducted on Cleveland data, but this data suffers from processing issues. In this research the best part is only two tests were analysed to perform the experiment, the tests that were considered

include lipid profile and ECG report, and other attributes don't require any specific test so reducing the time for test analysis.

3. Proposed Methodology

This part is based on techniques and the tools which are used to perform the comparison of the AI techniques for the diagnosis of cardiac disease.. The main steps followed are mentioned below.

- *Collection of data:* This is the first step in which information was gathered from the GMC Srinagar having just about 200 patients.
- *Pre-processing:* In this step the information was first arranged into the EXCEL to change over the record into CSV format then from CSV format to the ARFF document that is the format that weka tool can import.
- *Selection of the techniques:* This includes different AI techniques.
- *Training and testing:* There are various methods present but here split method was used in which 75% data are set to train the data and 25% and used for testing.
- *Summary analysis:* finally the results of Artificial Intelligence techniques implemented were compared on the basis of presence or absence of Triglycerides and Low – density lipoprotein.

Tools used

A. Excel

Excel is a spreadsheet application program used to record and analyse numerical and statistical data, distributed by Microsoft. It has number of amazing features which is used to perform various operations like calculations, pivot tables, graph tools, macro programming etc. Excel is also compatible with multiple OS like Windows, macOs, Android, and iOS. EXCEL supports the same operation, the data is maintained and manipulated in the groups of columns and rows. Not only this but the data can be presented by various interesting 2D and 3D graphs as well as with the histogram.

B. Weka

WEKA is one of the popular machine learning software that is written in java language. It an open source software which provides tools for data pre-processing, implementation and visualisation tools so that we can develop machine learning techniques and apply them to real- world data mining programs. It is developed in the country of New Zealand at the University of Waikato. Weka is under General Public License (GNU). The basic principal that is followed by the weka that the data should be present in the comma separated values only then the further processing can be done on this tool.

C. SPSS

SPSS is a Stands for statistical package for social science. This software was introduced by the IBM and is also known as IBM SPSS.SPSS tool can analyse and sort any kind of data. It can deal with almost all types of file formats like (.xls), (.doc), (.txt), comma separated VALUES.

4. Techniques Implemented

A. Artificial Neural Network

Artificial Neural Network is a network of input and output units where all the connections are associated with weights. The structure of AI is like the structure of a brain. It helps to build predictive models from large databases. It helps you to conduct image understanding, human learning, computer speech, etc. There are number of AI algorithms, most important is back propagation algorithm that became popular in 1980s. Back propagation method has been used in this paper. Back propagation in neural network is a short form for "backward propagation of errors." Learning of back propagation is based on the feed forward neural network. It is used by MLP for training. This algorithm works in an iterative manner, in this every iteration the output is compared with the desired one because error is calculated and back propagated to the network, this feedback then modifies the weight of neutrons and the finally we get the desired output on

Most prominent advantages of Back propagation are:

- It is fast, simple and easy to program
- It has no parameters to tune apart from the numbers of input
- It is a flexible method because does not require prior knowledge about the network

That paper describes several neural networks where back propagation works far faster than earlier approaches to learning, making it possible to use neural nets to solve problems which had previously been insoluble. Now the back propagation algorithm is the workhorse of learning in NN.

B. Decision Tree

Decision Tree is a tree like structure. It consists of several nodes and each node in the decision tree is attached to the decision tree through the edge that is path connecting two nodes. In medical domain the node represents the particular test and the results is represented by the corresponding edge [3] [4].Root node is also called the decision node that has the best predicting capability among all the existing data values then further splitting occurs on various parameters like information gain [22]. An interesting fact about the decision tree is that it works on the numerical data and categorical data, easy to implement and interpret [22].Decision Tree fallows the top-down approach while constructing tree that means starting from the decision node and reaching to the leaf node of the tree. It is a classification technique and with each input data value there is associated target value that is known as supervised learning [3]. The rules that are followed to build any decision tree is IF-THEN rules and the learning process in decision tree is also known as decision tree induction [22].There are number of decision-tree algorithms, some of algorithms are:

ID3: It is Iterative Dichotomiser 3.
C4.5: It is actually a successor of iterative dichotomiser.
CART: is a classification and regression tree.

There are two parameters which are involved in the construction of decision tree is entropy and information gain.

C. K mean clustering:

The concept of K-means came into existence in early 1950s but remains unpopular for a longer period of time until computing power became available. K means clustering is an exploratory data analysis technique that means it will explore the complete data set, implements non hierarchical method of grouping objects together that it will not follow any hierarchical method but it will take datasets as they are coming and will group them together. It is an unsupervised learning algorithm that belongs to the clustering technique.

Steps of K means:

- Input the number of cluster that we need to create whether we need to clutter in a size of two or a three so on so forth.
- After deciding the clusters calculate the centroid.
- After calculating the centroid, the distance between two centroid and clusters are created and they are clustered together on the basis of minimum distance.
- The process is repeated until we get the consistency in the cluster size and the consistency in the cluster assignment then we stop or till all the dataset have been completed.

Table 1: Primary Dataset of heart disease

Attribute	Type
Age	Continuous
Sex	Nominal
Blood Sugar	Continuous
LDL	Continuous
Triglycerides	Continuous
Cholesterol	Continuous
BP	Continuous
Chest Pain	Nominal
Resting ECG	Nominal
Diagnosis	Nominal

5. Experimental Result

As already discussed, the techniques like Decision Tree, Neural Networks were tested and trained on the below dataset using percentage and split method for first two techniques and classes to cluster evaluation for K-Means.

6. Conclusions

The paper enlightens the comparison of techniques Artificial Neural Network, Decision tree, K-means in the prediction of heart disease on a single dataset with two important parameters that is LDL and TG using machine learning set.

7. Future Work

Diagnosis of various types of heart diseases on the above mentioned parameters through machine learning and improvement of K means clustering in the diagnosis of cardiac disorders.

Table 2: Analysis of Techniques

Technique	Testing & Training Method (75% training and 25%)	Time taken to build the model		Significance
Multiplayer perceptron	Percentage split	92%	0.31sec	Highly significant among all techniques
Decision Tree	Percentage split	82%	0.02sec	Lesser significant than ANN
K Means	Classes to cluster evaluation	54%	0sec	Least significant

References

[1] E.P.Ephzibah1, Dr. V. Sundarapandian, A neuro fuzzy Expert system for heart disease diagnosis, an International Journal of Computer Science and Engineering Vol.2, No.1, February 2012.
[2] Vikas Chaurasia Saurabh Pal, Early Prediction of Heart Diseases Using Data Mining Techniques, Caribbean Journal of Science and Technology, 2013, Vol.1, 208-217.
[3] A. Sankari Karthiga, M. Safish Mary, M. Yogasins, Early Prediction of Heart Disease Using Decision Tree Algorithm, International Journal of Advanced Research in Basic Engineering Sciences and Technology, Vol.3, Issue 3, March 2017
[4] Mia Shouman, Tim Tuner, Rob stokerr "Applying K-nearest Neighbour in Diagnosing Heart Disease patients, International journal of Information and Education Technology, Vol 2, No 3, June 2012.
[5] Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin, Analysis of Data Mining Techniques for Heart Disease Prediction, International Conference on Electrical Engineering and Information and Communication Technology, IEEE 2016.
[6] Salma Banu N.K, Suma Swamy, Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey, 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimisation Techniques, IEEE 2016.
[7] Eesha Goel, Er. Abhilasha, Random Forest, International Journal of Advanced Research in Computer science and software Engineering, Volume 7, Issue 1, January 2017.
[8] Chi Zheng, Jingxin Liu, Tuberculosis Bacteria Detection based on Random Forest using Fluorescent Images, IEEE International Congress on Image and Signal Processing, 2016.
[9] Mohammad A. M. Abushariah, Assal et al, Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN)Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches, Journal of Software Engineering and Applications 2014.
[10] Miss. Manjusha B. Wadhonkar, Prof. P. A. Tijare, Prof. S. N. Sawlkar, "Classification of heart disease data set using multiple-layer feed forward back prolongation algorithm" International journal of Applications or innovation in engineering and management, Volume 2, Issue 4, April 2013.
[11] M. Akhiljabbar, B.L Deekshatulua Priti Chandra, Classification of Heart Disease Using K- Nearest

Neighbor and Genetic Algorithm, International Conference on Computational Intelligence: Modelling Techniques and Applications (CIMTA) 2013.
[12] Sheenal Patel, Hardik Patel, survey of data mining survey of data mining techniques used in health care domain, International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016.
[13] V. Manikantan, S. Latha, Predicting the analysis of heart disease symptoms using medicinal data mining methods, International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.
[14] G.Karthiga1, C.Preethi2, R. Delshi Howsalya Devi3, Heart Disease Analysis System Using Data Mining Techniques, International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2014.
[15] R. Tamilarasi, Dr. R. Porkodi, A Study and Analysis of Disease Prediction Techniques in Data Mining for Healthcare, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-3).
[16] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.
[17] R. Subha. k Anandakumar, A bharathi, “Stufy on cardia vascular Disease Classification Using Machine Learning Approaches, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016).
[18] Tülay Karay, lan Tülay Karay, lanv, Prediction of Heart Disease Using Neural Network, 2nd International Conference on Computer Science and Engineering (UBMK, 17), IEEE 2017.
[19] J. S. Sonawane and D. R. Patel, "Prediction of heart disease using-learning vector quantisation algorithm,” 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, pp.15.doi:10.1109/CSIBIG.2014.7056973.
[20] ESCAPE. (July 2017 February 2011) [Online] available http://www.unescap.org/stat//syb2009/9.
[21] Australian bureau of statistics. (July 2017- February 2011) [online-] available
[22] Anjikya, Nusrat af sheikh, “Data mining and visualisation for prediction of multiple diseases in health care”, International conference on Big Data analysis and Computational intelligence, IEEE2017.