Heart Disease Prediction using Multilayer Perceptron Algorithm

Article · May 2017

CITATIONS
14

READS
1,730

1 author:

Mudasir Kirmani
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir
68 PUBLICATIONS **328** CITATIONS

SEE PROFILE

ISSN No. 0976-5697

Volume 8, No. 5, May-June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Heart Disease Prediction using Multilayer Perceptron Algorithm

Mudasir M Kirmani
Assistant Professor, Computer Science
Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir
Srinagar, J&K, India

Abstract: Cardiovascular disease continues to claim an alarming number of lives across the globe. CVD disease is the greatest scourge affecting the industrialized nations. CVD not only strikes down a significant fraction of the population without warning but also causes prolonged suffering and disability in an even larger number. Although large proportion of CVDs is preventable they continue to rise mainly because preventive measures are inadequate. This paper presents the Multilayer Perceptron data mining technique to diagnose CVD. The discovered knowledge can be used by the healthcare administrators to improve the quality of service and also used by medical practitioners to reduce the number of adverse drug effect.

Keywords: Data Mining, Cardiovascular Disease (CVD), Disability Adjusted Life-Years (DALYs), low and middle income countries (LMICs), Multilayer Perceptron.

INTRODUCTION

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Cardiovascular disease continues to claim an alarming number of lives across the globe. CVD disease is the greatest scourge affecting the industrialized nations. CVD not only strikes down a significant fraction of the population without warning but also causes prolonged suffering and disability in an even larger number. In the united states alone CVD is responsible for almost one million fatalities each year and more than half of all deaths. Almost five million persons affiliated with CVD are hospitalized each year. The cost of these diseases in terms of human suffering and of material resources is almost incalculable. The epidemic of cardiovascular disease is a global phenomenon and in the current environment, the magnitude of this increase in incidence. Over the past decade CVD has emerged as the single most important cause of death worldwide. In 2010, CVD caused an estimated 16 million deaths and led to 293 million Disability Adjusted Life-Years (DALYs) lost accounting for approximately 30% of all deaths and 11% of DALYs lost that year. Now low and middle income countries (LMICs) are seeing an alarming and accelerating increase in CVD rates. [1]

Between 1990 and 2010, deaths from CVD increased from 26% to 29% of all deaths globally. Although large proportion of CVDs is preventable they continue to rise mainly because preventive measures are inadequate. Data mining plays an effective role for uncovering new trends in healthcare organization which is helpful for all the parties associated with this field. In order to improve the accuracy of diagnosis and to reduce the diagnosis time, an efficient and reliable Decision Support System for heart disease diagnosis using Multilayer Perceptron is developed. [2]

Literature Review: To diagnose cardiovascular disease using Multilayer Perceptron algorithm various research works have been done by different authors. They tried to achieve efficient methods and accuracies in finding out diseases related to heart by their work. Researchers have been investigating the use of data mining techniques to help practitioners in accurate diagnosis. Works done by various researchers in heart disease diagnosis using Multilayer Perceptron and other data mining techniques with various performance measures are discussed below:

Ms. Preeti Gupta and Ms. Punam Bajaj (2014) analyzed Multilayer Perceptron and swarm intelligence techniques for diagnosis of heart disease using heart rate data set. Swarm intelligence optimization i.e. Genetic Algorithm has been used and applied for optimizing the Multilayer Perceptron. [3]

Moloud Abdar, et al (2015) applied and compared data mining techniques to predict the rise of heart disease using five different algorithms such as C5.0, Naive Bayes, Support vector Machine, Multilayer Perceptron and Logistic Regression with accuracy measures as: 93.02, 86.05, 88.37, 85.22 and 80.23 using 13 medical attributes. [4]

G.Purusothaman and P.Kirshnakumari (2015) had cited various data mining prediction models namely Decision Table, Association Rule, KNN, Artificial Multilayer Perceptron, SVM, Naive Bayes and Hybrid models with accuracies as 76%, 55%, 58%, 85%, 86%, 69% and 96%. Hybrid data mining has outperformed other data mining heart disease diagnosing techniques. [5]

K.Aravintan and Dr. M. Vanitha (2016) did a comparative study on prediction of heart disease using cluster and rank based approach. Naive Bayes Algorithm, J48 Algorithm, Multilayer Perceptron Algorithm were used on 305 instances with 14 medical attributes. The accuracy measures of Naive Bayes, J48 and Artificial Multilayer Perceptron are as 81.3021%, 80.099%, and 82.56. [6]

Ashwini Shetty A and Chandra Naik (2016) analysed Multilayer Perceptron and Genetic Algorithm for heart disease prediction using 13 medical attributes with accuracy measures of 84% and 89%. [7]

K.Manimekalai (2016) did comparison of various data mining tools and techniques in which Naive Bayes gave 99.52% accuracy in the WEKA tool followed by Multilayer Perceptrons accuracy of 96.56% in .Net data mining tool and fuzzy logic's accuracy of 83.85% in Tanagra tool. [14]

Ankita R.Mokashi, et al (2016) analyzed improved K-Means and Artificial Multilayer Perceptron techniques for improving accuracy of heart disease with 13 number of medical profile parameters using 303 records of Cleveland database. As per results Artificial Multilayer Perceptron performed best. [9]

Basheer M. Al Maqaleh and Ahmad.M.G.Abdullah. (2016) developed a prototype for heart disease diagnosis using four experiments on data set of IBB hospital YEMEN and for all experiments two situations were considered one with 11 attributes and other with only 3 selected attributes. The accuracy measures with selected attributes of J48, Naive Bayes and Multilayer Perceptron were as 100%, 99.08% and 100%. [10]

Sellappan palaniappan et al. developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely Decision Tree, Naive Bayes and Multilayer Perceptron. Ihdps can answer complex "what if" queries which a traditional decision support system cannot. [11]

Vanisree k et al. Proposed a decision support system for diagnosis of congenital heart disease. The proposed system is designed and developed by using Mat lab's GUI feature with the implementation of back propagation Multilayer Perceptron. The proposed system achieved an accuracy of 90%. [12]

Dilip roy chowdury et al represent the use of Multilayer Perceptrons in predicting neonatal disease diagnosis. This study exhibits Multilayer Perceptron based prediction of neonatal disease and improves the diagnosis accuracy of 75% with higher stability. [13]

Artificial Multilayer Perceptron in data mining: A Multilayer Perceptron is a parallel, distributed information processing structure consisting of multiple number of processing elements called nodes; they are interconnected via unidirectional signal channels called connections. Multilayer Perceptron consists of one input layer, one output layer and one or more hidden layers. Each layer consists of one or more nodes, represented by small circles. The lines between nodes indicate flow of information from one node to another node. The input layer receives signals from external nodes. The output of input layer is given to hidden layer, through weighted connection links. It performs computations and transmits the result to output layer through weighted links. The output of hidden layer is forwarded to output layer, it performs computations and produce final result. The working of multilayer perceptron Multilayer Perceptron is summarized in steps as mentioned below and shown in figure 1:

1. Input data is provided to input layer for processing, which produces a predicted output.
2. The predicted output is subtracted from actual output and error value is calculated.
3. The network then uses a Back propagation algorithm which adjusts the weights.
4. For weights adjusting it starts from weights between output layer nodes and last hidden layer nodes and works backwards through network.
5. When back propagation is finished, the forwarding process starts again.
6. The process is repeated until the error between predicted and actual output is minimized.

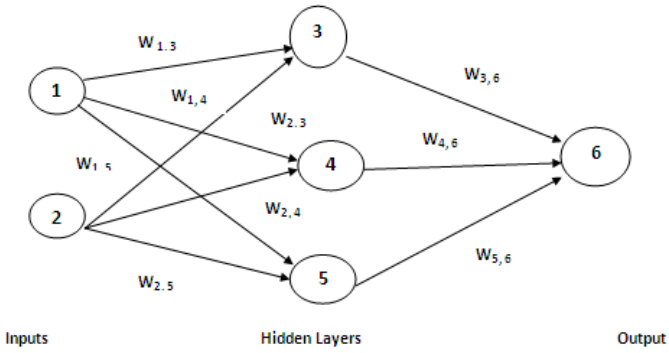

Figure 1: Artificial Multilayer Perceptron Model.

In this research paper, decision support system is developed for predicting heart disease of a patient. The prediction is done based on historical heart disease database.[14] The system uses medical terms such as sex, blood pressure, and cholesterol like 13 input attributes as shown in below given table:

Table 1: Description of Medical Attributes for Heart Disease Prediction

Sr. No	Attribute	Description	Values
1	Age	Age in years	Continuous
2	Sex	Male or Female	1=male 0=female
3	CP	Chest Pain Type	1=typical type 1 2=typical type angina 3=non angina pain 4=asymptomatic
4	Thestbps	Resting Blood Pressure	Continuous value in mm/hg
5	Cholestrol	Serum Cholestrol	Continuous value in mm/dl
6	RestECG	Resting Electographic Results	0= normal 1=having ST-T wave abnormal 2=left ventricular hypertropy
7	Fbs	Fasting Blood Sugar	1>=120 mg/dl 0<=120 mg/dl

8	Thalach	Maximum Heart Rate Achieved	Continuous value
9	Exang	Exercise Induced Angina	0= no 1= yes
10	Oldpeak	ST depression induced by exercise relative to rest	continous
11	Slope	Slope of the peak exercise ST segment	1= unsloping 2= flat 3=downsloping
12	Ca	Number of major vessels colored by floursopy	0-3 value
13	Thal	Defect type	3=normal 6=fixed 7=reversible defect

PERFORMANCE MEASURES

10 FOLD CROSS VALIDATION - In order to minimize the bias associated with the random sampling of the training and test data samples, 10-Fold Cross Validation was adopted. In 10-fold cross validation, the complete dataset is randomly split into 10 mutually exclusive subsets of approximately equal size. The classification model is trained and tested 10 times. Each time it is trained on nine folds and tested on the remaining single fold.

Confusion Matrix - In classification problems, the primary source of performance measurements is a confusion matrix. Given m classes, a confusion matrix is a table of at least size m by m as given in table 2.

Table 2: Confusion Matrix

	ACTUAL CLASS		
PREDICTED CLASS		**Class 1**	**Class 2**
	Class 1	TRUE POSITIVE	FALSE NEGATIVE
	Class 2	FALSE POSITIVE	TRUE NEGATIVE

The numbers along the diagonal from upper-left to lower-right represent the correct decisions made, and the numbers outside this diagonal represent the errors.

The overall accuracy of a classifier is estimated by dividing the total correctly classified positives and negatives by the total number of samples. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

RESULTS

The experiment is carried out on a publicly available dataset that consists of 303 instances. A data mining tool Weka 3.6.13 is used for experiment and the parameters used are listed in table 1. After applying Multilayer Perceptron on training dataset the results obtained is given in table 3:

Table 3: Results obtained after applying Multilayer Perceptron Model.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.824	0.21	0.824	0.824	0.824	0.894	<50
0.79	0.176	0.79	0.79	0.79	0.889	>50_1
0	0	0	0	0	?	>50_2
0	0	0	0	0	?	>50_3
0	0	0	0	0	?	>50_4
Weighted Avg. 0.809	0.194	0.809	0.809	0.809	0.891	

*Time taken to build model: 5.41 seconds

```
=== Confusion Matrix ===
   a   b   c   d   e   <-- classified as
 136  29   0   0   0 |   a = <50
  29 109   0   0   0 |   b = >50_1
   0   0   0   0   0 |   c = >50_2
   0   0   0   0   0 |   d = >50_3
   0   0   0   0   0 |   e = >50_4
```

out of 303 instances 136 cases are accurately predicted as true positive (that is number of persons who are correctly diagnosed having heart disease) ,29 instances are predicted as false negative(that is the number of patients who have heart disease but are shown as clear) 109 instances are exactly predicted as true negative (that is the no of cases who do not have heart disease and are exactly predicted as not having heart disease) and another 29 cases are wrongly predicted as false positives(that is person's not having heart disease are diagnosed as having heart disease).

CONCLUSION

The overall objective of this model, Heart disease prediction system (HDPS) is to predict accurately the presence of heart

disease using machine learning Multilayer Perceptron data miming algorithm. To validate the system 303 instances had been tested out of which 245 were correctly diagnosed with 80.8581 % and incorrectly classified instances are 58 that is 19.1419 %. The futuristic scope of the research work will be to build a hybrid system which can be used to diagnose heart disease prediction rate with higher levels of accuracy.

REFERENCES

1. http://www.world-heart-federation.org/press/fact-sheets/cardiovascular-disease-terms/. World Heart Federation, 2012.
2. Global Atlas on cardiovascular disease prevention and control Published by the World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, 2011.
3. Punam Bajaj, Preeti Gupta. Review on Heart Disease Diagnosis Based on Data Mining Techniques. International Journal of Science and Research.
4. Parvathi I, Siddharth Rautaray, Survey on Data Mining Techniques for The Diagnosis of Diseases In Medical Domain, International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 838-846.
5. G. Purusothaman and P. Krishnakumari. A Survey of Data Mining Techniques on Risk Prediction: Heart Disease, Indian Journal of Science and Technology, Volume 12, June 2015.
6. R. Subha, K. Anandakumar, A. Bharathi, Study on Cardiovascular Disease Classification Using Machine Learning Approaches. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4377-4380.
7. Ashwini Shetty, Chandra Naik, Different Data Mining Approaches for Predicting Heart Disease. International Journal of Innovative Research in Science, Engineering and Technology. Vol. 5, Special Issue 9, May 2016.
8. K.Manimekalai , A Proficient Heart Disease Prediction Method Using Different Data Mining Tools. Volume 6 Issue No. 3 International Journal of Engineering Science and Computing, March 2016.
9. Ankita R. Mokashi, Madhuri N. Tambe, Pooja T. Walke. Heart Disease Prediction Using ANN and Improved K-Means. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol. 4, Issue 4, April 2016.
10. Basheer Mohammed Al-Maqaleh, Ahmed Mohammed Gasem Abdullah, An Intelligent and Electronic System based Classification and Prediction for Heart Disease Diagnosis. International Journal of Emerging Trends in Science and Technology.
11. Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Technique, 978-1-4244-1968-5/08/$25.00 ©2008 IEEE.
12. Vanisree K, Jyothi Singaraju, Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks, International Journal of Computer Applications (0975 – 8887) Volume 19– No.6, April 2011. (50)
13. Dilip Roy Chowdhury, Mridula Chatterjee & R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011.
14. Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease