STM JOURNALS

Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720
Volume 6, Issue 3
www.stmjournals.com

Heart Disease Diagnosis Using Artificial Neural Networks

Hilal Ahmad Bawani[1,*], Amjad Husain[2]
[1]Student, Department of Information Technology, Central University of Kashmir, Ganderbal, Jammu and Kashmir, India
[2]Assistant Professor, Department of Information Technology, Central University of Kashmir, Ganderbal, Jammu and Kashmir, India

Abstract

Coronary illness is a term covering any defect of the heart. In contrast to cardiovascular problem, which portrays issues with the veins and circulatory framework just as the heart, coronary illness alludes to issues and deformations in the heart itself. As indicated by the Centers for Disease Control (CDC), coronary illness is the main source of death in the United Kingdom, United States, Canada, and Australia. One of four passing's in US happens because of coronary illness. It is a treatable malady whenever analyzed at opportune time; however, that was less conceivable by customary way. So, it needs a superior and effective approach to for determination at beginning period. In this paper, a heart disease prediction system based on neural networks with 14 clinical medical features as input to check whether a person suffers from heart disease or not with high accuracy.

Keywords: *Cleveland dataset, neural networks, multilayer perceptron, deep learning*

****Author for Correspondence*** E-mail: hilalbawani414@gmail.com

INTRODUCTION

As per the reports of the World Health Organization (WHO), heart disease is killer disease around the world [1]. About seventeen million people suffered life loss due to this deadly disease during a single year. But, it can be cured if diagnosed at right stage but that is very much difficult by using traditional systems.

Coronary illness is a general term that implies that the heart is not working properly. Children can be born into the world with coronary illness. This is called inherent coronary illness. On the off chance that individuals get coronary illness later, it is called obtained coronary illness. There is always a demand to develop such an automated system that will help in diagnosis of heart disease with high accuracy and efficiency. So, it has been proposed that an automated system for diagnosis of heart disease in this paper based on machine learning.

Upcoming sections will cover all insights, proposed methodologies, proposed system, and conclusion.

LITERATURE SURVEY

Numbers of researchers have contributed their efforts to develop a basis for heart disease field. Deep learning have played a vital role in detecting the heart disease at right stages, nowadays research is going on how to improve the efficiency and accuracy of such systems.

There are a lot of datasets available but many of the researchers have told that the majority of datasets are not providing such information that will result in high accuracy and efficiency. It has been found that Cleveland dataset is among best datasets that is being used by majority of the researchers.

One of the tedious works is to extract crucial features from dataset as it simplifies the training process and reduces the complexity in terms of time. Jones et al. were first to develop a heart disease diagnosis system based on neural networks using a self-applied questionnaire. This study covers overall data about this disease including risk factors. This proposed was validated against three risk factors ("also called Dundree Rank Factor Score") that include blood pressure, cholesterol and smoking. They have used multilayered neural network based on an algorithm namely backward propagation [2].

Haung et al. have used the method of vector quantization algorithm, and they have used

thirteen clinical features. They carried out a three stage process, in first stage of this process they identified 13 important features and during second stage they use a classifier and in last stage they developed a system and found its accuracy around eighty percent [3].

Sonawane and Patil used another type of methodology for this disease prediction. They used VQA with random incremental training. They selected 13 features as input to neural network and in hidden layer they used the variable number of nodes at hidden layer and at output layer they used a single node. It was found that the accuracy of this model was around 85% [4].

Another methodology by Feshki and Shijani for heart disease prediction, they used neural network as classifier and their methodology were split into three steps. In first step they simply split dataset into two parts, one part included sick and other part included healthy. In second step 8192 subsets have been extracted and was done out of total features. Third step, the subset with more accuracy was found by PSO algorithm. They used four classifiers for this and the accuracy was found around 91% [5].

Another approach was proposed by Manza et al. for heart disease diagnosis, they used neural network as classifier and their methodology were split into five steps. In first step, they collected data about medicines prescribed by doctor to patient. In second step, data are converted to machine understandable format. In third step, conversion function is used to process data given on input layer. In step four, classifier performance was evaluated [6].

Amin et al. have proposed a model that is combination of various methods and artificial neural network. The biggest disadvantages of neural network are finding the starting (initial) weights and convergence problem. The accuracy of this model Wion system is a as around 90% [7].

Bashir et al. have proposed a model based on decision tree, SVM and Naïve bayes algorithms. They just calculate the output of these algorithms and then use the combined form of it. The accuracy of this model was 82% but has got an improved specificity of 93% [8].

METHODOLOGY

As far as heart disease prediction is concerned, the majority of the researchers have used the artificial neural networks (ANNs) as it is the only methodology applicable for nonlinear relationships among data. This heart disease prediction system is a supervised type of learning.

Artificial Neural Networks

ANN learning is powerful against blunders in the preparation information and has been effectively connected for finding genuine value, discrete-value, and vector-value capacities containing issues, for example, translating visual scenes, discourse acknowledgment, and learning robot control systems. The investigation of fake neural systems (ANNs) has been propelled to some degree by the perception that natural learning frameworks are worked of complex snare of interconnected neurons in cerebrums. The human cerebrum contains a thickness between associated systems of roughly 10^{11}–10^{12} neurons; each associated neuron by and large associated to $l0^4$–10^5 other neurons. So normal human mind take in exact 10^-1 to settle on shockingly complex choices. ANN framework is motivated to catch this sort of exceptionally parallel calculation dependent on appropriated portrayals.

For the most part, ANNs are worked out of a thickly interconnected arrangement of straightforward units, where every unit takes various genuine esteemed sources of info and produces a solitary genuine esteemed yield. However, ANNs are less spurred by natural neural frameworks; there are numerous complexities to organic neural frameworks that are not displayed. There are three layers, one is input layer, the hidden layer where processing occurs and the output layer. Input layer is the layer where a user provides its input. Hidden layer is the layer where actual processing of data takes place [9]. Output layer is the layer where the output is available after processing. The idea of ANNs is based on the belief that working of human brain by

making the right connections can be imitated using silicon wires living neurons and dendrites. The human brain is composed of 86 billion nerve cells called neurons. They are connected to other thousand cells by Axons. Stimuli from external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural network. A neuron can then send the message to other neuron to handle the issue or does not send it forward. ANNs are composed of multiple nodes, and imitate biological neurons of human brain.

The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value. Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values.

Backpropagation Neural Network

Backpropagation is a directed learning calculation, for preparing multilayer perceptron (artificial neural networks). While structuring a neural network, at the outset, we introduce weights with some irregular qualities or any factor for that reality. Presently clearly, we are not superhuman. In this way, it is redundant that whatever weight esteems we have chosen will be right, or it accommodates our model the best. We have chosen some weight esteems first and foremost, yet our model yield route is not quite the same as our real yield for example, the mistake worth is immense. Presently, by what method will we decrease the mistake? Fundamentally, what we have to do, we have to some way or another disclose the model to change the parameters (loads), with the end goal that blunder ends up least. We should place it in another manner. One approach to prepare our model is called as backpropagation.

The weights are first instated to irregular number that is generally little. A preparing capacity that takes input information at information layer and procedures it and gives yield, this capacity is called sigmoid capacity here.

$$\sigma=\sigma\ w.x\ (\ ) =1/1+e-y$$

Where, w represent weight and x represent input value.

Error function is calculated for output node and is propagated to all input nodes similarly propagated to hidden layer nodes as:

$$\delta k \leftarrow ok\ 1-ok\ (tk-ok)$$

Where, sigma represents error function of output unit k and o(k) represents network output for k. Similarly sigma function calculated for hidden nodes. After that weights gets updated unless and until this error functions tends to zero.

EXPERIMENTAL RESULTS

This system was developed in Jupiter notebook and it uses multilayer neural network.

Data Source

In this, we have used Cleveland dataset as it was very popular among researchers. It consists of 303 instances and 76 attributes but majority of scholars used only 14 features among these. We have created a dummy values for attribute chest pain in order to categorize it to get clear results about a patient. We have used 80% data to train and 20% data to test in order to well train this model and get high accuracy (Table 1).

Table 1: *Features and Their Description.*

Features	Description
Age	Represents Age
Sex	Defines gender male=1, female=0
Cp	Level of pain in chest
trestbps	blood pressure at rest
chol	cholesterol (serum)
fbs	fasting blood sugar> 120 mg/dl (1 = true; 0 = false)
restecg	resting electrocardiographic results
thalach	Maximum heart rate reached
exang	angina induced by exercise
oldpeak	induced by exercise ST depression
Slope	ST segment slope
Ca	Major vessels count
Thal	Thal level values
Num	Decision making feature

Performance Evaluation

Accuracy can be calculated as ratio of number of correct predictions by total predictions Accuracy=(TP+TN)/(TP+TN+FP+FN)

Where, TP is true positive that is patient having heart disease is correctly diagnosed with disease, TN is true negative that is patient not having disease is declared healthy by system correctly, FP are reverse of TP and FN is reverse of TN.

Recall represents actual proportion having disease. Recall=TP/(TP+FN). Precision represent instances that actually have heart disease and are diagnosed with it. Precision=TP/(TP+FP). But it is necessary to focus on such instances that will result in better accuracy and precision. So we have to first select the features and prune the hidden network layer size (Tables 2 and 3). The reason of the hidden layer nodes must be reduced in order to improve accuracy and precision by PCA [5].

Table 2: *Accuracy and Precision Dependency on ANN Hidden Layer Count.*

Hidden layer size	Accuracy (%)	Recall (%)	Precision (%)
3	82.22	85.71	78.26
4	75.55	77.77	66.66
5	84.44	86.82	89.28
6	75.55	85.18	76.66
7	82.22	84.00	84.00
8	86.66	86.95	86.95
9	71.11	73.91	70.83
10	86.66	89.47	80.95
11	77.77	80.95	73.91
12	84.44	86.95	83.33

Table 3: *Optimization by PCA.*

Hidden layer size	Accuracy (%)	Recall (%)	Precision (%)
3	91.11	84.61	100.00
4	88.88	95.45	84.00
5	88.88	88.88	92.30
6	86.66	89.47	80.95
7	91.33	100.00	89.00
8	90.55	95.45	95.00
9	91.11	95.83	88.46
10	91.11	100.00	85.10
11	91.11	100.00	91.66
12	90.55	95.00	88.00

CONCLUSION

The system proposed for heart disease prediction is a multilayer neural network that uses backpropagation algorithm and uses 22 input nodes and only one output node. It basically uses 14 features of Cleveland dataset and after processing those it comes to a conclusion whether a person is sick or not. This proposed system will provide an accuracy of above 90% with good Precision and Recall.

REFERENCES

1. New initiative launched to tackle cardiovascular disease, world's number one killer, World Health Organization [online]. Available from http://www.who.int/cardiovascular_diseases/global_hearts_initiative/en/ [Accessed on July 2019].
2. Shen Z, Clarke M, Jones RW, Alberti T. *Detecting the risk factors of coronary heart disease by use of neural networks*, *Proceedings of 15th Annual International Conference of IEEE Engineering in Medicine and Biology Societ*, 1993, 277–288p. doi:10.1109/IEMBS.1993.97841.
3. Chen AH, Huang SY, Hong PS, Cheng CH, Lin EJ. HDPS: Heart disease prediction system, 18-21 Sept. *2011 Computing in Cardiology*, Hangzhou, China.
4. Sonawane JS, Patil DR. *Prediction of heart disease using learning vector quantization algorithm*, 2014, *Conference on IT in Business, Industry and Government (CSIBIG)*, Indore, 2014, 1p. doi:10.1109/CSIBIG.2014.7056973
5. Feshki MG, Shijani OS. Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network, 9 April 2016 Artificial Intelligence and Robotics (IRANOPEN), 48–53p. Qazvin, Iran. doi: 10.1109/RIOS.2016.7529489
6. Hannan SA, Mane AV, Manza RR, Ramteke RJ. *Prediction of heart disease medical prescription using radial basis function*, 2010, *IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, 2010, 1–6p. doi: 10.1109/ICCIC.2010.5705900

7. Amin SU, Agarwal K, Beg R. *Genetic neural network-based data mining in prediction of heart disease using risk factors,* 2013, *IEEE Conference on Information & Communication Technologies*, JeJu Island, 2013, 1231p. doi:10.1109/CICT.2013.6558288
8. Bashir S, Qamar U, Younus Javed M. *An ensemble-based decision support framework for intelligent heart disease diagnosis, International Conference on Information Society (i-Society 2014),* London, 2014, 259–264p. doi: 10.1109/i-Society.2014.7009056
9. Tulay Karayilan, Ozkan Kilic, *Prediction of heart disease Using Neural Network,* 2017, *IEEE conference on computer science and engineering.*

Cite this Article
Hilal Ahmad Bawani, Amjad Husain. Heart Disease Diagnosis Using Artificial Neural Networks. *Journal of Artificial Intelligence Research & Advances*. 2019; 6(3): 32–36p.

