JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HEAT FAILURE PREDICTION WITH NEURAL NETWORK USING DESIGN THINKING FRAMEWORK

Jagadesh M / Assistant Professor
Department of Electronics and Communication Engineering
SNS College of Technology
Coimbatore,India

Dharshana R / UG Student
Department of Electronics and Communication Engineering
SNS College of Technology
Coimbatore,India

Darunya S / UG Student
Department of Electronics and Communication Engineering
SNS College of Technology
Coimbatore,India

Ahiesh Sujo I S / UG Student
Department of Electronics and Communication Engineering
SNS College of Technology
Coimbatore,India

Dhilip R M / UG Student
Department of Electronics and Communication Engineering
SNS College of Technology
Coimbatore,India

Abstract* — *Machine learning is frequently used to teach machines to learn and predict, when given with new data. A person's risk of developing heart illness may be predicted in a number of ways thanks to advancements in machine learning. One of the illnesses that affects individuals the most in the entire globe is heart disease. Heart disease is a risk that can be increased by a number of causes. By merging logistic regression analysis with neural networks, a novel method of heart disease prediction is presented. In order to determine the key risk factors for disease prediction, logistic regression is first performed. It creates substantial risk factors that might be utilized to anticipate cardiac disease based on statistical p-value. Identifying and removing risk factors with little or no impact is done. The neural network receives the important emerging elements. Additionally, a neural network trained on the risk factors determined from logistic regression is employed to identify whether or not a person has heart disease. As a result, heart disease is predicted using a mix of logistic regression and neural networks.

***Keywords*—** *Logistic regression model, neural networks, prediction, heart disease*

I. Introduction

Heart disease is one of the most common conditions in the world. The proper functioning of the heart is critical to the body. Heart ails include myocardial infarction, myocardial ischemia, natural heart complaint, coronary heart complaint, cardiac arrest, supplemental heart complaint, and others. Predicting cardiac complaint may be done using a variety of styles.

Machine learning is a kind of artificial intelligence in which machines learn from training data and make prognostications on test data grounded on what they've learnt. The core principle underpinning machine literacy is to identify patterns in data and induce prognostications. Machine literacy has several uses, including recommender systems, medical diagnostics, bioinformatics, and so on.

In machine learning, there are three forms of literacy supervised literacy, unsupervised literacy, and underpinning literacy. Prophetic analytics encompasses a variety of statistical approaches similar as prophetic modelling, machine literacy (ML), and data mining that are used to produce prognostications grounded on current or once data. Prophetic analytics is used in client relationship operation, the healthcare sector, and numerous other diligences. Deep learning has a large influence on prophetic analytics. Prophetic modelling includes multitudinous models similar as Naive Bayes, Logistic retrogression, Neural networks, Support Vector Machine, Bracket and Retrogression trees, and so on. One fine or computational fashion is the artificial neural network (ANN). It's similar to the neurons in the mortal brain. There are connections, propagation directions, and separate layers in the artificial neural network. Each subcaste is comprised of bumps with arrows representing the connections between them. The input subcaste of a neural network has multitudinous bumps. These input subcaste bumps are linked to the bumps in the retired subcaste. Weights are allocated to each input. The network's input bumps give data to the retired subcaste, which conducts colorful tasks or computations and sends the reused data to the affair knot. The knot that produces the final result is located in the affair subcaste. This is an overview of the neural network process.

II. LITERATURE SURVEY

In this section, we briefly review several related works on the heart failure prediction,

[1] V. Sree Hari Rao, et al., "Novel Approaches for predicting Risk Factors of Atherosclerosis," states that coronary heart disease (CHD) caused by cholesterol-induced atherosclerosis kills many people worldwide. Slow, asymptomatic disease development can cause cardiac arrest, stroke, or myocardial infarction. Imaging is used to predict risk by understanding atherosclerotic plaque molecular and metabolic activities. Imaging technologies can provide some plaque metabolism information but lack the resolution and sensitivity for identification. This paper predicts CHD risk factors based on clinical observations and behaviors.

[2] Paolo Melillo, et al., "Classification Tree for Risk Assessment in Patients Suffering from Congestive Heart Failure via Long-Term Heart Rate Variability", states that, this research will automatically classify congestive heart failure patients' risk (CHF). Patients are risk-classified using long-term HRV assessments. NYHA defines patients as low-risk or high-risk (NYHA). Two public Holter databases retrospectively analyzed 12 moderate (NYHA I and II) and 32 severe (NYHA III and IV) CHF patients. Patients with NN/RR ratios over 80% were eligible to ensure signal quality. Created CART classifiers. 30 high-risk and 11 low-risk patients were examined. Classification trees identified high-risk patients with 93.3% sensitivity and 63.6% specificity. Finally, CART's principles are simple and complement prior research indicating decreased HRV might estimate CHF risk.

[3] Minas A. Karaolis, et al., "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining with Decision Trees," states that CHD kills and disables most people in wealthy countries. CHD identification and therapy need more investigation. Data mining found cardiac event risk factors to minimise CHD incidence. Before: age, sex, and family history of premature CHD; after: smoking, systolic and diastolic blood pressure, total cholesterol, HDL, LDL, triglycerides, and glucose. MI, PCI, CABG investigated (CABG). Age, smoking, and hypertension were the biggest risk factors for MI, PCI, and CABG. Other studies obtained most risk factors. MI, PCI, and CABG models had the highest correct categorization rates: 66%, 75%, and 75%.

[4] Carlos Ordonez, "Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction", states that, associations may predict cardiac disease. Unfortunately, medical data sets generate many association rules. These guidelines are mostly medically unimportant and difficult to locate. Association rules are frequently mined without independent sample confirmation. We propose using search limitations to reduce the number of rules, searching for association rules on a training set, and validating them on an independent test set. Support, confidence, and lift determine medical relevance of regulations. Association restrictions apply to heart disease patient medical records. Risk factors and cardiac perfusion are linked to four-artery disease. Search constraints and test set validation reduce association rules and provide predictive rules. Multiple runs on the test set validate important rules with high confidence, lift, or both. Important medical guidelines.

III. EXISTING SYSTEM

The Existing system for heart disease prediction is to have a regular check-up of the heart's working and then meeting the doctor for feedback and how to maintain the health.

The following are some of the system's drawbacks:

a. Time Consuming as the doctor may be in any emergency cases
b. Human fault can occur.

IV. PROPOSED SYSTEM

The ultimate objective is to forecast cardiac disease using both the neural network-based technique and the logistic regression model. The heart disease dataset consists of 303 individual observations, of which 297 observations are taken into account. The suggested system is composed mostly of two components. The first step is to use the available risk variables in the dataset and the p-value to identify the significant risk factors for predicting heart disease. This p-value produces the important codes for each attribute. The dataset is then split into a training dataset and a testing dataset as the second step. The neural network was created for the training dataset, and it can predict the testing dataset after learning new information.

A. Data Gathering

The Cleveland Heart Disease database is where the information for this research was sourced. 297 data with 14 medical characteristics used to forecast heart disease. The dataset's summary is listed in Table.

S.NO	ATTRIBUTE NAME	DESCRIPTION
1.	Age	Age in years
2.	Sex	1= male, 0= female
3.	CP	Chest pain type (1= typical angina, 2= atypical angina, 3= non-anginal pain, 4= asymptomatic
4.	TrestBps	Resting blood pressure (in mm Hg on admission to hospital)
5.	Chol	Serum Cholesterol in mg/dl
6.	Fbs	Fasting blood sugar >120 mg/dl (1= true, 0= false)
7.	RestecG	Resting electrographic results (0= normal, 1= having ST-T wave abnormality, 2= left ventricular hypertrophy
8.	Exang	Exercise-induced angina
9.	Oldpeak	ST depression induced by exercise relative to rest
10.	Slope	Slope of the peak exercise ST segment (1= upsloping, 2= flat, 3= downsloping)
11.	Ca	Number of blood vessels colored by fluoroscopy
12.	Thal	3= normal, 6= fixed defect, 7= reversible effect
13.	Num	Class (0= healthy, 1= have heart disease)
14.	Thalach	Maximum heart rate

B. Logistic Regression Model

One statistical regression model that can quantify the connection between a categorical dependent variable and one or more independent variables is the logistic regression model. Age, sex, the type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrographic results, maximum heart rate, exercise-induced angina, old peak-slope of the peak, exercise-induced peak slope, blood vessels affected, and that defect are the independent variables in this case.

The class that is to be identified as either having heart disease or being in good health is the dependent variable. The dependent variable in a binary logistic regression analysis is coded as "0" or "1" depending on whether the subject has heart disease or not.

The logistic regression model calculates the likelihood of having the disease based on the risk factors. The conditional probability p(y=1 | X) can be calculated, where X = (x1, x2...xn) denotes the disease's n risk factors. As a result, we were able to determine how likely someone would be to develop the condition. You can choose 0.5 as the cutoff value. The individual is thought to have heart disease if the cutoff value is greater than 0.5; otherwise, the individual is seen to be heart disease-free. In addition, the logistic regression model has the ability to choose elements based on the statistical significance p-value that have a substantial impact on heart disease.

The probability value, or p-value, is expressed in shorthand as the probability and is indicated by the summary of the logistic regression model. P-values are employed in statistical hypothesis testing, which is used in various academic domains like political science, economics, etc. The marginal significance used to describe the likelihood that a particular event will occur is known as the p-value.

Out of the fourteen qualities in the dataset, the logistic regression model can identify the key risk factors or characteristics that are utilized to predict heart disease. When using the logistic regression model, the risk factors or characteristics whose pvalues are less than 0.05 (p0.05) are included as variables.

The statistically significant p-value is higher than 0.1 (p>0.1) for variable exclusion criterion. The greater pvalue shows that there is no correlation between changes in the independent variable and changes in the dependent variable. The important risk factors identified by the logistic regression model's statistically significant p-value include sex, the nature of the chest discomfort, resting blood pressure, fasting blood sugar, exercise-induced angina, the slope of the peak exercise, blood vessel damage, and thalamic defect.

C. Dataset for Training and Testing

The 297 records in the dataset [16] are split into a training dataset and a testing dataset. The training dataset is the collection of instances used for learning and fitting the classifier's weights, and it is used to create a predictive connection. The test set is a collection of samples used to gauge how well a fully stated classifier performs. The training dataset is split into 25% and 25% of the testing dataset, respectively.

D. Building Neural Network

A computational model based on biological neural networks is known as a neural network. The foundation of artificial neural networks (ANN) is the study of the human brain. The human brain is a highly intricate network of neurons [17]. An analogous ANN is a network containing three interconnected components, such as input, hidden, and output. The risk factors or characteristics of the patient are inputs utilized in medical diagnosis.

In the field of medicine, artificial neural networks have proven to be effective. In order to forecast coronary heart disease, ANN are employed. Here, the 8 neurons that make up the input layer correlate to 8 important characteristics. One output class variable accepts either a value of 0 or 1.

A number of 0 indicates that the person does not have heart disease, whereas a value of 1 indicates that they do. Three nodes are utilized in the hidden layer. Fig. 1 displays the sample artificial neural network. High precision is a key benefit of neural networks. Applications for neural networks include fraud detection, medical [14], and accountancy. The neural network can predict if there will be heart disease in the testing dataset based on the learned network or training dataset.

E. Performance Measures

The accuracy, specificity, and sensitivity of neural networks are used to determine several performance metrics. The neural network achieved an accuracy of 84 percent. The obtained sensitivity and specificity are 91.4 percent and 77.5 percent, respectively.

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

$$\text{Specificity} = TN / (TN+FP)$$

$$\text{Sensitivity} = TP / (TP+FN)$$

where,

TP = True Positive; that is the number of samples which are classified as having heart disease while they were actually have heart disease.

TN = True Negative; that is the number of samples which are classified as not having heart disease while they were actually not.

FN = False Negative; that is the number of samples which are classified as not having heart disease while they were actually have heart disease

FP = False Positive; that is the number of samples which are classified as having heart disease while they were actually not.

The accuracy, sensitivity and specificity of Neural Network is tabulated in Table 2.

classifier	Accuracy	Sensitivity	Specificity
Neural Network	84%	91.4 %	77.5 %

V. ADVANTAGE OF PROPOSED SYSTEM

The advantages of the proposed systems are listed below:

- Less need of papers
- Can check it remotely

VI. CONCLUSION AND FUTURE SCOPE

Machine learning is an active research topic, with many researchers working in the healthcare domain to identify illness risk. The interpretability of model parameters and ease of use are advantages of logistic regression. The benefit of neural networks is that they can discover complicated non-linear correlations between dependent and independent variables with no conventional statistical training. The combination of logistic regression with neural networks results in a revolutionary way to forecasting an individual's heart disease. The future work may be extended for longitudinal investigations of the patients and to increase the accuracy in prediction of heart disease.

VII. REFERENCES

[1] V. Sree Hari Rao, M. Naresh Kumar, "Novel Approaches for predicting Risk Factors of Atherosclerosis," IEEE Journal of Biomedical and Health Informatics., vol 17, No.1, Jan 2013.

[2] Paolo Melillo, Nicola De Luca, Marcello Bracale and Leandro Pecchia, "Classification Tree for Risk Assessment in Patients Suffering from Congestive Heart Failure via Long-Term Heart Rate Variability", IEEE Journal of Biomedical and Health Informatics., Vol. 17, No. 3, May 2013.

[3]Minas A. Karaolis,Joseph A. Moutiris, Demetra Hadjipanayi, Constantinos S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees," IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 3, May 2010

[4] Carlos Ordonez ,"Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction", IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, April 2006.

[5] Sikander Singh Khurl, Gurpreet Singh, "Ranking Early Signs of Coronary Heart Disease Among Indian Patients", IEEE International Conference on Computing for Sustainable Global Development, 2015

[6]The UCI Machine Learning Repository[online] https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[7]Stephenie C. Lemon, Jason Roy, and Melissa A. Clark, Peter D. Friedmann, William Rakowski, "Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison with Logistic Regression", The Society of Behavioral Medicine, Vol. 26, No. 3, 2003, pp. 172-181.

[8]David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant, "Applied Logistic Regression", 2013.

[9]Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction ", International Journal of Computer Applications, Vol. 17, No.8, March 2011, pp. 43-48.

[10]Harleen Kaur, Siri Krishan Wasan and Vasudha Bhatnagar, "The Impact of Data Mining Techniques on Medical Diagnostics", Data Science Journal, Vol. 5, October 2006, pp. 119-126.

[11] K.Srinivas, B.Kavihta Rani , A.Govrdhan , "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering (IJCSE), Vol. 02, No. 02, 2010, pp. 250-255.

[12]Asha Rajkumar, G.Sophia Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm", Global Journal of Computer Science and Technology, Vol. 10 , September 2010.

[13]Qeethara Kadhim Al-Shayea, "Artificial Neural Networks in Medical Diagnosis ", International Journal of Computer Science Issues (IJCSI), s, Vol. 8, No. 2, March 2011, pp. 150-154.

[14]R. Dybowski and V. Gant, "Clinical Applications of Artificial Neural Networks", Cambridge University Press, 2007.

[15]Tutut Herawan, Rozaida Ghazali, Nazri Mohd Nawi, Mustafa Mat Deris, "Recent Advances on Soft Computing and Data Mining", The Second International Conference on Soft Computing and data mining (SCDM-2016).