Efficient Heart Disease Prediction for Large Volume of data using ANN

Rohith Ravi[1]
Department of Computer Science and Engineering
SRM Institute of Science and Tech- nology
Chennai, India

Ganesan R[2]
Assistant Professor School of Computing
SRM Institute of Science and Tech- nology
Chennai, India

Yadlapalli Bodha Bhargav[3]
Department of Computer Science and Engineering
SRM Institute of Science and Tech- nology
Chennai, India

***Abstract*—**In this model we are going to develop an efficient prediction model(system) of and for Heart Disease using neural networks. With today's modern lifestyle, large populations of the world are getting diagnosed with heart diseases. It is becoming one of the topmost reasons for loss of life. As the death rate is ever increasing as a result of coronary heart diseases, the people working close with healthcare depend largely on the patient's data to predict if the patient may have a risk of heart disease. Due to several underlying contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnor- mal pulse rate and many other factors, hence, it makes heart disease prediction challenging. Various methods in data mining and neural networks have been implemented to explore the austerity of heart disease like Artificial Neural Network (ANN), Regression Algorithm, K -Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), Naive Bayes (NB) and others.

The aim of this model is to find best predicting algorithm which can help the general and inexperienced doctors or medical technicians in predicting the risk of the heart disease. The prediction system we are proposing uses ANN with back propagation and by using artificial neural network along with back propagation a better accuracy can be achieved. The diagnosis of heart disease in most cases depends on a complex combina- tion and collection of huge volume of clinical and pathological data. We will be using the following data, 1. Age 2. Sex 3. Blood Pressure 4. Heart Rate 5. Diabetes 6. Hyper cholesterol 7. Body Mass Index (obesity) as the probable input. Furthermore, Machine learning has been shown to be effective and efficient in assisting in making decisions and predictions from the large quantity of data produced and circulated by the Health Care Industry.

Therefore, our proposed prediction model is going to be a life saver as it will be able to provide accurate diagnosis for heart problem susceptible patients and hence the respective treatment can be undergone as quickly as possible and improve the patient's life condition and expectancy considerably.

Keywords—Prediction; Machine Learning; ANN; Back Propagation; Attributes;

I. INTRODUCTION

Many patients of today ranging from toddlers to elderly people go to hospitals and clinics with concerns and complains of chest pains and pains relating to both the right and left chest. With them having no clue what the issue is, this is where disease prediction comes in to the picture and filtering it to Heart Disease prediction by assessing the regions of pain and conducting some vital tests. The results from the test like the cholesterol level, bp, and the ECG is not enough to exactly predict what Heart condition or disease the patient has. It is difficult to identify the exact specific heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors.

Hence for an efficient and accurate Prediction of Heart Dis- ease, the use of Machine Learning Algorithms plays a big role through using algorithms like KNN, RNN and we are proposing to use ANN with back propagation which yields an accuracy of above 90% in the role of predicting and di- agnosing what exact Heart Disease or condition the patient has. The diagnosis of heart disease in most cases depends on a complex combination and huge volume of clinical and pathological data. Machine learning has been shown to be effective assisting in making decisions and predictions from the large quantity of data produced by the health care indus- try. It is very important to take care of our health to make sure that we don't get ourselves into a predicament in the future.

II. RELATED WORK

A. Prediction of Heart Disease Using Machine Learning

This is the 1st paper we went through which was written by Aditi Gavhane; Gouthami Kokkula; Isha Pandya; Kailas Devadkar. The author had looked at an existing system which uses 76 parameters to detect whether a person has a heart disease. Some of the parameters included expensive clinical tests like ECG, CT scan etc. which are very costly. So, they considered taking 7 important parameters which are
1. Age 2. Sex 3. Blood Pressure 4. Heart Rate 5. Diabetes 6. Hyper cholesterol 7. Body Mass Index (obesity). They also considered finding all factors using simple sensors which are in watches, phones. They used 4 sensor detectors which were 1. AliveKor 2. MyHeart 3. HealthGear 4. Fitbit
They detect factors like Indexes like Height, Weight, BMI - Blood Pressure, Haemoglobin, WBC, RBC, Platelets - [Lipids]: Cholesterol, HDL, LDL, VLDL, triglycerides, - [Sugar]: Fasting Glucose after meals, HbA1C, pulse rate, BP, calories burned. Now in the proposed system MLP is used to train and test the dataset.

B. Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification

The author [3] has proposed a hybrid method which im-plements various kinds of techniques and hopefully on fur-ther development can be implemented in such a way that it can achieve almost 100% accuracy. In the 1st step they fed normalized data to the network and the expected result is achieved. The actual and computed result is checked for error differences. Error, weights and bias are rearranged. In case the error exceeds the tolerance then step1 is again fol-lowed else process is terminated. Data Mining Techniques used are 1) Neural Networks 2) Decision Trees 3) Naive Bayes 4) Bayesian Network 5) Support Vector Machine. They also said that Neural Networks and Naive Bayes tech-nique perform the best among the above Data Mining tech-niques and b/w Neural Networks and Naive Bayes the most efficient way to proceed is using Neural Networks. The al-gorithm used is Back propagation to achieve the best results.

C. Prediction of Heart Disease Using Neural Network

The author has proposed a technique which uses Back propagation algorithm using ANN learning methodology. The input variables are disease risk factors which are ob-tained from datasets. The Dataset taken by the author is the most commonly used dataset in heart disease prediction which is Cleveland Database present in UCI. The input lay- er was designed in such a way that it contains 13 neurons which was equal to the number of attributes in the data set. The Hidden layer was developed in a way that it had 3 neu- rons and the o/p. Proposed heart disease prediction system has been designed as a Multilayer Perceptron Neural Net- work. The Layer was designed in such a way that it had 2 layers. Proposed heart disease prediction system has been designed as a Multilayer Perceptron Neural Network

D. Prediction of Heart Disease Using Learning Vector Quantization Algorithm

The author used multilayer perceptron in this system. The system has 2 essential steps which were accepting 13 clinical data [4] as input and as a last step training the net- work by Back-propagation Algorithm.

E. Prediction of Heart Disease Medical Prescription Using Radial Basis Function

The authors main goal is to predict any irregularities of heart diseases by using an ANN as a classifier. The process had five steps. In stage one data related to pharmaceuticals for heart diseases was collected, then in stage two, abnor-malities in heart diseases and medicines were converted to binary characters. 1 indicates that a medicine or abnormality is present. In stage 3, Radial Basis Function was finally used for training the model. Then in stage 4, testing data is per- tained to verify the execution of the classifier. Then in stage 5, Radial Basis Function is appointed to determine the pharmaceuticals for valetudinarian or patient. A lot of hid- den layers are present in the Radial Basis Function units. An accuracy of 97% [5] can be achieved by employing this method. This method can be improved by using Generalized Regression Neural Network.

F. Efficient heart disease prediction system using decision tree

The authors have used clinical data to detect any kind of abnormalities in the cardio-vascular system. They have used the Cleveland database which is used very commonly for heart disease prediction. They have used many 76 raw at- tributes [6] in detection but for the final result they use only 14 to detect it. They have used a tool called KEEL (Knowl- edge Extraction Based on Evolutionary Learning). It is an open-source Java software tool that can be used for a large number of different knowledge data discovery tasks. KEEL allows to perform a complete analysis of new computational intelligence proposals in comparison to existing ones.

G. Analytical study of heart disease diagnosis using classification techniques

The author considers using Apriori algorithm and SVM (support vector machine) in heart disease prediction. Using attributes such as age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict the likelihood of patients getting heart diseases. From the research the author has proved that classification-based techniques contribute high effectiveness and obtain high accuracy compared to other methods. Apriori algorithm uses prior knowledge of fre- quent item-set properties. We apply an iterative approach or level wise search where k-frequent item sets are used to find k+1 item-sets. This algorithm is generally used for finding frequent item-sets in a dataset for Boolean association rule.

H. ELECTRONIC RECORDING SYSTEM-HEART DISEASE PREDICTION SYSTEM

The authors here have considered using an automated que-tionnaire like decision support and application system for data collection, which further helps in the prediction and diagnosis of a possible vulnerable heart disease. The predic-tion application system is to be done using Naive Bayesian Classification technique and the diagnostic and remedy gen-eration uses Pattern Matching Algorithm. The data collec-tion includes attributes like the patient's health history, com- plete clinical data and their appointment details as well all in a single database for ease of use and access. After the re-spective data has been entered and collected, data mining comes into place to figure out the probability if the patient is vulnerable to heart disease or not. If the presence of heart disease is positive, then the symptoms entered by the patient are compared with the database using Pattern Matching Al-gorithms and then based on this the remedy is generated. The main purpose and intention of the proposed model and system is to reduce costs through avoiding overtime pay and at the same time increases the number of patients who can be accurately treated.

I. Design And Implementing Heart Disease Prediction Using Naives Bayesian

The author has used data mining to measure all extricating of very critical, concealed and important data from huge information bases. Typically, the Healthcare area includes bountiful of information identified with patients, different analysis of the illnesses and so forth These days the medical clinics are receiving the way of life of emergency clinic IMS (data the executives' frameworks) to handle their or patient's information efficiently and adequately. Enormous amount of information is created by such frameworks that is spoken to utilizing graphs, numbers, text and pictures. However, such kind of information is not really utilized for making any clinical choices. The ebb and flow research accentuates on coronary illness finding. Different strategies of information mining have been joined for diagnosing the sickness in this manner gettinga few probabilities. Concern- ing the coronary illness expectation various frameworks are being suggested which are being sent by the methods for different strategies and calculations. Increasing quality assistance at reasonable cost remains the prime and testing worry for the medical care foundations. For offering quality administrations at standard, there must be precise finding of the patients alongside viable dose of drugs. Bad quality clin- ical determination and treatment can yield in undesired and lacking outcomes. One answer for cost cutting by Health- care foundations can be usage of PC created information or utilization of DSS (choice help frameworks). Normally the Healthcare area includes plentiful of information identified with patients, different conclusion of the sicknesses, asset the board and so on This data or information must be further separated by the Human administrations.

J. Predicting Heart Disease at Early Stages using Machine LEARNING : A Survey

The authors here have emphasized on the usefulness of the detection and diagnosis of heart diseases at an early stage in order to prevent further casualties and the severity. The fact that Machine Learning Algorithms have proven to be effec- tive in recent times with heart disease prediction has been brought to effect. The supervised Machine Learning Algo- rithms applied by the authors in the prediction here are arti- ficial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes) (NB) and knearest neighbour algorithm. Furthermore, the several types of Heart Diseases have been elaborated speci- fying the most common ones being Coronary Heart Disease (CAD) and Heart Failure (HF) and the causes behind them brought up as well. From the survey, the different risk fac- tors leading to Heart Failure and Coronary Heart Disease have been mentioned specifically and they are mainly of two types, controllable and uncontrollable. The several methods to diagnose heart diseases have been told and An- giography being the most used trendy method, due to its drawbacks and limitations in recent times, automated sys- tems using ML have been encouraged to more use due to its analytical capability using large sets of collected clinical data. Some of the attributes used in Prediction are Gender, Age, FBS, RBS, Sex and types of Chest pain, Heart Rate, ECG Results and etc.

III. Model Diagram

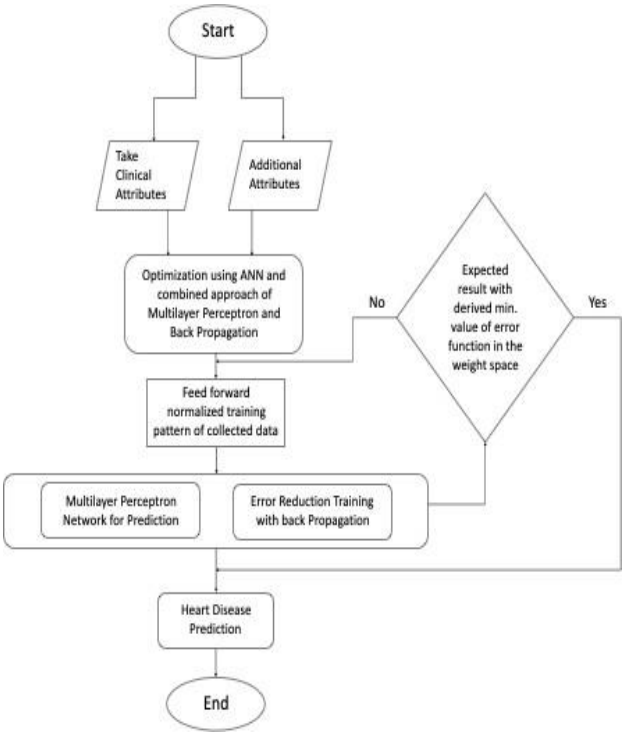

Figure1: Flow chart for ANN working

IV. Inference from the Survey

Our goal in the model is to develop a working model which can predict whether a person has any heart disease by exploring different parameters, and determine whether the person has any existing threat to any kind of pulmonary disease. We have found out a few parameters which might help us out like age, sex, CP (Constrictive pericarditis (**CP**) is a form of diastolic **heart** failure that arises because an inelastic pericardium inhibits **cardiac** filling. This disorder must be considered in the differential diagnosis for unex- plained **heart** failure, particularly when the left ventricular ejection fraction is preserved), Trestbps (**trestbps**: resting blood pressure (in mm Hg on admission to the hospital) chol: serum cholestoral in mg/dl. fbs: fasting blood sugar > 120 mg/dl (1 = true; 0 = false) restecg: resting electrocar- diographic results), chol (**Cholesterol** is a waxy substance found in your blood. Your body needs **cholesterol** to build healthy cells, but high levels of **cholesterol** can increase your risk of heart disease. With high **cholesterol**, you can develop fatty deposits in your blood vessels.), fbss (Fasting Blood Sugar: A blood glucose test measures the amount of a type of sugar, called glucose, in your blood. A few different types of blood glucose tests are used. Fasting blood sugar (**FBS**) measures blood glucose after you have not eaten for at least 8 hours. It is often the first test done to check for prediabetes and diabetes.), restecg (resting electrocardio-

graphic. Results(restecg)), thalach (maximum heart rate achieved), exang (exercise induced angina), oldpeak (ST depression induced by exercise relative to rest), slope

(exer- cise ST segment), ça (number of major vessels), thal (tha- lassemia).

We can also consider adding additional factors like smoking and any activity which increases the chance in heart diseases can be added.

We also took some time into knowing the different kinds of heart diseases which are Angina, Arrhythmia, Fibrillation, Congenital heart disease, Coronary artery disease, Myocar- dial infarction, Heart failure.

V. FACTORS

We can use various algorithms to detect a Cardiovascu- lar disease. The algorithms used are ANN, CNN, Data Min- ing, Trees, SVM's, Evolutionary Rule Learning technique and various other methods to determine whether the person has the disease or not. So normally a person who smokes a lot will have a higher chance in getting a disease like aortic aneurysm. Smoking causes narrowing of vessels and it may occur to atherosclerosis in the blood vessels. The nicotine in s m o k e: L e s Cardiovascular e n s t h e a m o u n t o f o x y g e n your heart acquires, uprears your BP, paces up your heart beat, Increases the chances of acquiring blood clots more, this may cause heart attacks/strokes, it harms the interior part of your blood vessels, encompassing the ones even in your heart. Also having a bad diet may lead upto a heart disease at times. So, diet may be included in as a factor for determine in the prediction of a coronary disease. Diet is an important factor in coronary heart diseases. Food-related risk elements include obesity, high BP, diabetes and a diet high in saturated fats. A low-fat diet will decrease the chance s in getting a coronary disease so it is important to eat nutritious food. So, obesity can be added as a factor in determining whether the person will get a coronary disease or not.

RISK FACTORS WHICH AREN'T IN OUR CONTROL	RISK FACTORS WHICH ARE IN OUR CONTROL
Age	Smoking status
Gender	Diet
Ethnicity	Cholesterol Levels
Family History	Blood pressure
	Body weight
	Diabetes (Sugar Management)

Figure-II Factors determining Cardio Vascular Diseases

VI. IMPLEMENTATION

1) Data Pre-Processing : -

In the model we created we used Anaconda IDE which uses python and for the GUI we used Tkinter which is the basic Graphic user interface for python. We used ANN with back propagation which will fine tune the weights of a neu- ral net based on the error rate obtained in the previous epoch(period). Proper adjustment of the weights permits you to diminish error rates and to make the model attested by escalating its generalization. Backpropagation is a short configuration for "backward propagation of errors." It is a quality system in training artificial neural networks. This procedure assists in calculating the gradient of a loss func- tion with respect to all the weights in the network. It is a viable method.

How Backpropagation Works: Simple Algorithm

Inputs X, come from the pre-connected track.

Step One is to input the model using real weights. Weights are frequently by random selection then we com- pute the o/p for every neuron from the input layer, to the hidden layers, then finally to the output layer. Then we compute the bias in the outputs.

ErrorB= Actual Output – Desired Output

Equation-(1)

Then we travel back from the output layer to the hidden layer to attune the weights such that the bias is eliminated or minimized.

We keep repeating this process until we are satisfied with the final outcome. Few people have predicted that with the help of Back propagation 100% accuracy is achievable and an accuracy of 90% is achievable with ease.

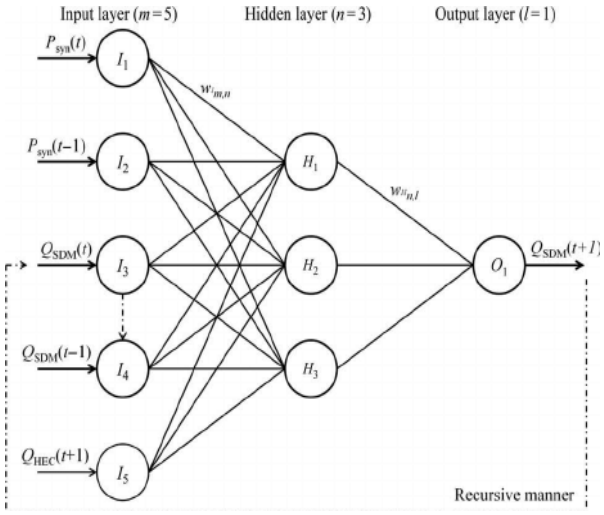

Figure-III Back Propagation Architecture

Most pre-eminent advantages of Backpropagation are: Backpropagation is quick, straightforward to code or program. It has no parameters to tune apart from the numbers of input. It is a pliable method as it doesn't need preliminary command or mastery in the field.

It is a quality method that usually works well.

It doesn't require any exceptional broach of the attributes of the function to be competent in.

This is one of the main reasons why we went with back-propagation, because it is very easy to learn and one of the most efficient ways to do it.

So, to sum up the whole process in a few steps: I)Initializing the Network. (Pre-processing) II)Forward Propagate.

III)Detect the Back Propagate Error. IV)Train Network.

V) Predict.

VI) Heart Disease Dataset Case Study.

VII. Objective

The objective of the model is to provide a means of accurately and quickly predicting the heart disease the patient has been diagnosed with. This will help considerably in sav- ing a lot of lives at the earliest as possible and prevent from further complications leading to and from the possible heart disease. The objective of the model is to develop a working model which can predict whether a person has any heart disease by looking at different parameters, and determine whether the person has any existing threat to any kind of pulmonary disease. Also, one of the objectives in the work- ing model is to take the maximum number of parameters as it helps a lot in increasing the possibility in detecting the disease. The motto is to decrease the cost if the patient isn't in particular rush and is willing to lay back for a while. If there shows any kind of symptoms the patient will need to consult a doctor as soon as possible. Especially if you are aged between 40 and 75 years old and have never had a heart attack or stroke, the model can roughly calculate the chance of having a coronary disease in the next 10 years.

REFERENCES

1. A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technolo- gy (ICECA), Coimbatore, 2018, doi: 10.1109/ICECA.2018.8474922.
2. A. Dewan and M. Sharma, "Prediction of heart disease using a hybrid technique in data mining classification," 2015 2nd International Con- ference on Computing for Sustainable Global Development (INDIA- Com), New Delhi, 2015.
3. T. Karayılan and Ö. Kılıç, "Prediction of heart disease using neural network," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017 , doi: 10 . 1109 / UBMK.2017.8093512.
4. J. S. Sonawane and D. R. Patil, "Prediction of heart disease using learning vector quantization algorithm," 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, doi: 10.1109/CSIBIG.2014.7056973.
5. S. A. Hannan, A. V. Mane, R. R. Manza and R. J. Ramteke, "Predic- tion of heart disease medical prescription using radial basis function," 2010 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2010, doi:

10.1109/ ICCIC.2010.5705900.

6. Purushottam, K. Saxena and R. Sharma, "Efficient heart disease pre- diction system using decision tree," International Conference on Computing, Communication & Automation, Noida, 2015, doi: 10.1109/CCAA.2015.7148346.
7. C. Sowmiya and P. Sumitra, "Analytical study of heart disease diag- nosis using classification techniques," 2017 IEEE International Con- ference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputhur, 2017, doi: 10.1109/ ITCOSP.2017.8303115.
8. S. Shaikh, A. Sawant, S. Paradkar and K. Patil, "Electronic recording system-heart disease prediction system," 2015 International Confer- ence on Technologies for Sustainable Development (ICTSD), Mum- bai, 2015, doi: 10.1109/ICTSD.2015.7095854.
9. A. N. Repaka, S. D. Ravikanti and R. G. Franklin, "Design and Im- plementing Heart Disease Prediction Using Naives Bayesian," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, doi: 10.1109/ICOEI.2019.8862604.
10. R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, doi: 10.1109/ICESC48915.2020.9155586.