Biomedical Application of Predicting a Health Of Patient

[1]Gauri Vinayak Sonawane, [2]Dr.V.S.Narayana Tinnaluri
[1]Research Scholar, [2]Head of the Department
[1]Sandip University,
[1]Sandip University, Nashik, India

Abstract : Analyze different parameters of patient .Heart disease and high blood pressure is deadly disease that large population of people around the world suffers from .When considering death rates and large number of people who suffers from heart disease. Traditional way of diagnosis is not sufficient for such an illness. Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way. A heart disease prediction system which uses artificial neural network backpropogation algorithm is proposed. various cinical features were used as input for the neural network and then the neural network was trained with backpropogation algorithm to predict absence or presence of heart disease with accuracy of 97%.also we will provide notification to the doctor, nurse and any medical staff on their handset

IndexTerms **- -Heart disease, artificial neural network, Cleveland database, backpropogation algorithm, multilayer perceptron, machine learning.**

I. INTRODUCTION

Peoples who suffer from the heart disease is the one of the killer disease according to world health organization (WHO) statistics. Predicting symptoms of heart disease early plays a crucial role for the treatment. If heart condition of the patients could be predicted before, lots of patient deaths would prevented and also a more accurate and efficient treatment way could be provided. Developing a medical diagnosis system regarding machine learning for prediction of heart disease provides more accurate diagnosis than traditional way and reduces cost of treatment. Predicting the heart disease by an automated medical diagnosis system based on machine learning is proposed to satisfy this need. Backpropogation algorithm which is commonly used artificial neural Network learning methodology, we used number of assumption of predicting the heart disease.

II. Related Work

A.Artificial Neural Network(ANN) takes an inspiration from human brain which has incredible processing ability because of having webs of interconnected neurons. The problem which are not linearly separable problem should be solved by multilayer perceptron neural network. The designed ANN has three layers: namely an input layer, an output layer and hidden layer.

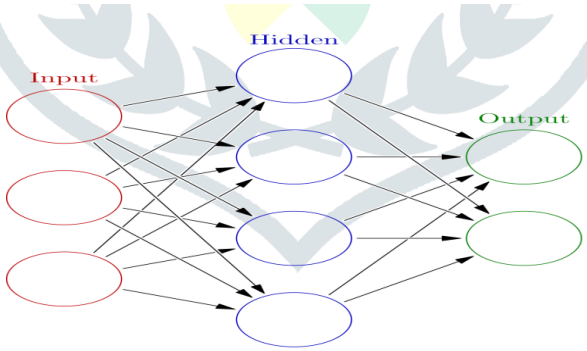

Fig 1.ANN architecture

2.1 Input layer was designed to contain 13 neurons. Number of neurons was decided to be equal to the numerous of attributes in the data set.

b.Hidden layer was designed to contain 3 neurons. This number was decided as a startup point. The number was changed increasing one by one until it reached to the neurons of hidden layer should be the mean of the numerous of the neurons of hidden layer should be the mean of the numerous of the neurons of input layer and output layers.

c. Output layer was designed to contain 2 neurons. The designed NN is a classifier going running in machine mode which means returning a class label (e.g., "Disease presence"/"Disease Absence"). Deciding 2 neurons is based on idea that the output layer consist one node per class label in model.

2.2. Backpropogation Algorithm:

Backpropogation algorithm (BA) is the most usually used ANN learning technique. The steps of the algorithm are listed below:

- All network weights are initialized to minimum random numbers.
- Training data is received as input and output is calculated for each unit with equation below known as sigmoid function:

$$O=o(w \parallel x) \qquad o(y)=1/1+e^{-y}\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where w is a vector of unit weight values and x is vector of network input values.

- Then error computational step is started. BP algorithms works as follows: Error signal(σ) which is calculated for each network output is propogated to all neurons in the network as input.
- Error term Sigma h is calculated for each network. output is propagation unit *k* using following equation: Sigma h←Oh (1-oh) ∑wkh Sigma k

where Wkh denotes network weight from hidden unit *h* to output unit *k*.

- Each network weight is updated where

Wij←wij+Delta wji where delta Wij=n sigma j*xij

where ή is learning rate and denotes the input from unit *i* into unit *j*. Backpropagation Algorithm was used for the proposed system as learning algorithm. 13 of the attributes of Cleveland dataset was used as input data for the designed neural network. The dataset was split into three parts: training, testing and validation. Then training was done with Backpropagation Algorithm.

2.3. Performance Evaluation:

The progress of the proposed system was computed by different metrics like accuracy, precision, recall. Accuracy is computed dividing number of predictions which are correct by number of all predictions. The obtained result is multiplied by hundred to get value as percentage.

$$Accuracy= (TP+TN) / (TP+TN+FP+FN)$$

Where TP, TN, FP, and FN demonstrate in order of the no. of true positives, true negatives, false +ve and -ve. TP demonstrates the number of instances which are sick and diagnosed accurately. FP demonstrates the no. of instances which are healthy and diagnosed wrongly are they are sick. FN demonstrates the no. of instances which are sick but the instances are diagnosed wrongly. TN contains a number of instances which are healthy and the instances are diagnosed accurately. Precision denotes the ratio of the instances that are predicted as having heart disease actually have heart disease.

$$Precision =TP/(TP+FP)$$

Recall denotes the proportion of the instance that are actually have heart diseases are predicted as having heart disease.

$$Recall=TP/(TP+FN)$$

Accuracy, recall and precision were decided to express success of predicting heart disease system. Using only accuracy can be sometimes misleading.

Sometimes selecting a model in which has lower accuracy is desirable , because it provides more about robust predictor for the problem . All predictions can be predicted as the value of majority class by model, when problem domain has a large class imbalance. This model is not useful when considered problem domain. This is referred to namely Accuracy Paradox. For such problems classifiers should be evaluated with additional measures.

To improve performance, dimensionality reduction with Principal Component Analysis (PCA) was done by reducing number of neurons of the input layer from 13 neurons to 8 neurons . The results which are obtained by changing hidden layer size with reduced dimensionality are seen in table A below.

Hidden Layer Size	Accuracy	Recall	Precision
3	91.111111%	84.615385%	100.00000%
4	88.888889%	95.454545%	84.000000%
5	88.888889%	88.888889%	92.307692%
6	86.666667%	89.473684%	80.952381%
7	93.333333%	100.00000%	89.285714%
8	95.555556%	95.454545%	95.454545%
9	91.111111%	95.833333%	88.461538%
10	91.111111%	100.00000%	85.185185%
11	95.555556%	100.00000%	91.666667%
12	91.111111%	95.652174%	88.000000%

III. DATA MINING TECHNIQUES

Data mining techniques are used to in order to learn, analyze and extract medical using many different algorithms in order to discover unknown patterns. Researchers data mining techniques for the analyze of many diseases such as heart disease, diabetes, stroke and cancer and many data mining techniques have been used in the diagnosis of heart disease with good accuracy. Researchers have been applying different data mining techniques such as naïve Bayes, neural network, decision tree, bagging,

kernel density, and support vector machine for prediction and diagnosis of heart diseases. On of the systems uses neural based learning classifier for classifying data mining tasks showed that neural based learning classifier system performs equivalently to supervise learning classifier. IEHPS intelligent and effective heart attack prediction system was built using data mining and neural networks and it proposed extracting significant patterns for heart disease prediction using K-means clustering and used MAFIA algorithm to mine the frequent patterns. Polateral developed system using hybrid fuzzy and K-nearest neighbor approach for the prediction of heart disease, which had 87% accuracy in diagnosis. After that we will provide notification to the doctor, nurse about heart condition of patient on to their mobile. After all this record doctor will be monitor health of patient this is smart way to analyze patient health condition. This project very much important in rural and urban area.

IV.METHODOLOGY

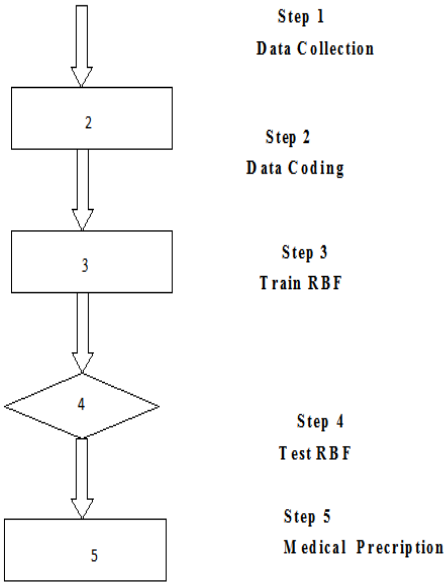

Fig 2. Flow chart of RBF

i Step 1 is collection of heart disease patient and the medicines provided by the doctor. The detailed information about the prescribed by the doctor is the output.
ii In step 2 , we have converted the symptoms of heart disease and medicine prescribed by the doctor is converted in to binary form i.e. either 1 or 0. If the symptom or medicine is present means it is 1 and if the symptoms are absent it is 0. Each numeric data is presented with one digit only i.e. either 0 or 1.
iii Step 3 is train RBF, we set up the RBF parameters including input, hidden and output layer neuron number, learning cycles and so forth.
iv Step 4, we apply testing data to evaluate the performance of the trained RBF.
v In step 5, medicines prescribed by the for the heart disease patient.

The data is collected from Civil hospital , Nashik while heart specialist examining the patients. The symptoms and information about patients details like Previous History(p1), present history(p2), Personnel History(p3), Physical Level examination(p4), Cardio vascular system(CVS) , Respiratory Rate(RS), Per Abdomen(PA) , Central nervous system(CVS), ECG and Blood Investigation(BI). The main point is ECG from which patient can easily diagnose whether the patient is having heart problem or not. All 200 patients' data collected regarding heart disease and the data are prepared in different Excel sheets which contains codes of each individual disease, history and symptoms. In one excel file 13 sub-sheets are taken for each field of information such as for Previous History(p1), for Present History the second sub-sheet and the name is given(p2), for Personnel History(p3) the third sub-sheet is taken, like this data collection has 13 different sub-sheets for different fields. All the fields are taken under the supervision of the cardiologist. The code is given to each symptoms, physical level examination parameter or diseases in each sub-sheet for experimental work. On this data some pre-processing i.e. normalization, coding and decoding methods are applied for the expected output.

The previous history has represented with 1 to 18 different diseases of total 300 heart patients and represents the codes respectively from 1to 18. The code 1 which represents Hypertension, code 2 represents Diabetes Mellitus like this it contains 18 different diseases which have codes from 1 to 18. The present history and the symptoms present in p2 are presented by codes. The code 1which represents Chest Pain/ Discomfort, code 2 represents Retrosternal pain like this it contains 29 different symptoms is defined by its corresponding code. Like this we have different 10 sheets, sheet 3 contains personnel history which contains 4 codes, sheet 4 contains physical examination which contains 25 codes , sheet 5 contains 8 codes . In sheet 11 all the medicines names long with their codes i.e. MID which is prescribed by the doctor to the patient. The medicine contains 52 different medicines 300 stages of the heart disease patient.On a Android phone doctor and nurse will see heart condition of patient.also web base application will generated..

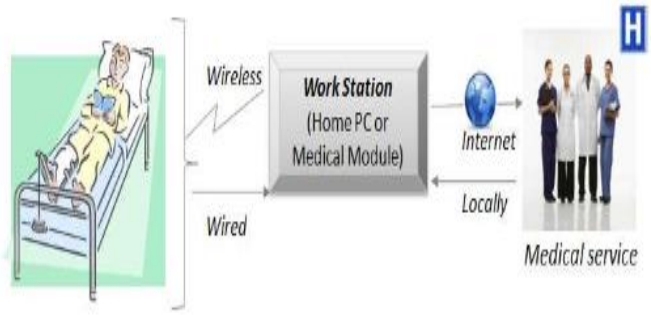

Fig 3. Typical E-health Monitoring System

Clinical Features	Description
Age	Age
Ca	Number of major vessels (0-3) colored by fluoroscopy
Chol(mg/dl)	Serum cholesterol
Cp	Chest pain type
Exang	Exercise induced angina
Fbs	Fasting blood sugar
Oldpeak	ST depression induced by exerciser elative to rest
Restecg	Resting electrocardiographic results
Sex	Gender
Slope	The slope of the peak exercise ST segment
Thal	3=normal ; 6 = fixed defect; 7= reversible defect
Thalach	Maximum heart rate achieved
Trestbps(mmHg)	Resting Blood Pressure

Table 1.Clinical features and descriptions

V. CONCLUSION

The proposed heart disease predictive system has been designed as a many layer perceptron neural network. It was Trained with backpropogation algorithm in order to predict whether heart disease present or not in the patient . There are a lot of studies on prediction of heart disease. Results of these studies vary up to almost accuracy of 100% . The proposed system gives 97% accuracy rates which means a very good rate according to related studies on this field . As further study, the proposed methodology can be enhanced as a hybrid model with other classification algorithms in order to obtain more accurate diagnosis for heart disese.

REFERENCES

[1]. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009 Feb 19;457(7232):1012–1014. doi: 10.1038/nature07634. PubMed CrossRef

[2]. White R, Horvitz E. From web search to healthcare utilization: privacy-sensitive studies from mobile data. J Am Med Inform Assoc. 2013 Jan 1;20(1):61–68. doi: 10.1136/amiajnl-2011-000765. http://jamia.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=22661560. PMC free article PubMed CrossRef

[3]. White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc. 2014;21(1):49–55. doi: 10.1136/amiajnl-2012-001473. http://jamia.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=23666794. PMC free article PubMed CrossRef

[4]. Horvitz E, Mulligan D. Machine learning and inference makes it increasingly difficult for individuals to understand what others can. Science. 2015;349(6245):253–255. http://research.microsoft.com/en-us/um/people/horvitz/science-2015-horvitz-mulligan-253-5.pdf. PubMed

[5]. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc. 2006:244–248. http://europepmc.org/abstract/MED/17238340. PMC free article PubMed

[6] Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J Oncol Pract. 2016 Jun 7;:737–744. doi: 10.1200/JOP.2015.010504. PubMed CrossRef

[7]. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc. 2013 May 1;20(3):404–408. doi: 10.1136/amiajnl-2012-001482. http://jamia.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=23467469. PMC free article PubMed CrossRef

[8]. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res. 2013;15(6):e124. doi: 10.2196/jmir.2614. http://www.jmir.org/2013/6/e124/ PMC free article PubMed CrossRef

[9]. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav. 2006 Aug;9(4):494–498. doi: 10.1089/cpb.2006.9.494. PubMed CrossRef

[10]. Hu Z, Hao S, Jin B, Shin A, Zhu C, Huang M, Wang Y, Zheng L, Dai D, Culver DS, Alfreds ST, Rogow T, Stearns F, Sylvester KG, Widen E, Ling X. Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. J Med Internet Res. 2015;17(9):e219. doi: 10.2196/jmir.4976. http://www.jmir.org/2015/9/e219/ PMC free article PubMed CrossRef

[11]. Rakovski C, Rosen A, Wang F, Berlowitz D. Predicting elderly at risk of increased future healthcare use: How much does diagnostic information add to prior utilization? Heal Serv Outcomes Res Methodol. 2002;3(3-4):267–277. doi: 10.1023/A:1025866331616. CrossRef

[12]. Phithakkitnukoon S, Smoreda Z, Olivier P. Socio-geography of human mobility: a study using longitudinal mobile phone data. PLoS One. 2012;7(6):1–9. doi: 10.1371/journal.pone.0039253. http://dx.plos.org/10.1371/journal.pone.0039253. PMC free article PubMed CrossRef

[13]. Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. Circulation. 2011 Mar 15;123(10):1116–1124. doi: 10.1161/CIRCULATIONAHA.110.943860. http://circ.ahajournals.org/cgi/pmidlookup?view=long&pmid=21403122. PMC free article PubMed CrossRef

[14]. Wilkinson D. Knowing when to stop: futility in the intensive care unit. Curr Opin Anaesthesiol. 2011;24(2):160–165. doi: 10.1097/ACO.0b013e328343c5af. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3252683/ PMC free article PubMed CrossRef

[15]. Yan J, Liu N, Wang G, Zhang W. How much can behavioral targeting help online advertising?. Proceedings of the 18th International Conference on World Wide Web; 18th International Conference on World Wide Web; Apr 20-24, 2009; Madrid. 2009. pp. 261–270. CrossRef

[16]. Chen J, Stallaert J. An economic analysis of online advertising using behavioral targeting. MIS Quarterly. 2010;38(2):429–449. http://ssrn.com/abstract=1787608.

[17]. Barr C, Jones R, Regelson M. The linguistic structure of English web-search queries. Proceedings of the Conference on Empirical Methods in Natural Language Processing; Empirical Methods in Natural Language Processing; Oct 25-27, 2008; Honolulu. 2008. pp. 1021–1030. CrossRef

[18]. Jadhav A, Andrews D, Fiksdal A, Kumbamu A, McCormick JB, Misitano A, Nelsen L, Ryu E, Sheth A, Wu S, Pathak J. Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. J Med Internet Res. 2014;16(7):e160. doi: 10.2196/jmir.3186. http://www.jmir.org/2014/7/e160/ PMC free article PubMed CrossRef

[19]. Callahan A, Pernek I, Stiglic G, Leskovec J, Strasberg HR, Shah NH. Analyzing information seeking and drug-safety alert response by health care professionals as new methods for surveillance. J Med Internet Res. 2015;17(8):e204. doi: 10.2196/jmir.4427. http://www.jmir.org/2015/8/e204/ PMC free article PubMed CrossRef

[20]. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst. 2009 Nov 01;27(4):1–37. doi: 10.1145/1629096.1629101. CrossRef