Artificial Neural Network for Prediction of Breast Cancer

Parag Singhal
Computer Science Department
Engineering College Bikaner
Bikaner(Rajasthan),INDIA
Paragsinghal63@gmail.com

Saurav Pareek
Computer Science Department
Engineering College Bikaner
Bikaner(Rajasthan),INDIA
Sauravpareek02@gmail.com

Abstract:

Technologies play a vital role in cancer care. Data mining approach has helped a lot in medical science due to its high efficiency in the prediction of the future health condition, and also helps in reduction of medical cost and improving the health of people and quality in real time which helps in saving lives' of people.Breast Tissue is the reason due to which breast cancer develops. It takes the second place for the most cancer diagnoses in the women after skin cancer. There are humungous data and information which gives an opportunity for analysing and understanding the process and make some researches using machine learning techniques. The main part of this paper is to create a tool for early prediction of breast cancer with the highest accuracy possible and low error rate. This was done by applying machine learning algorithms and with help of Artificial Neural Network (ANN) using Wisconsin Breast Cancer (Diagnostic) Dataset. Experimental results show that ANN gives the accuracy up to98% with low error rate. The Experiment is conducted using Dev.-C++ software and implemented using C-language.

Key words: Breast Cancer, ANN, Back-Propagation.

I. Introduction

Breast cancer is the most common cancer in women in India and accounts for 27% of all cancers in women [1, 2]. During 2012, in India, 144,937 were new cases for women for development of breast cancer in which 70,218 women died of breast cancer so if we calculate 144937 / 70218 = 2.06 which rounds off to 2. So roughly, for every 2 women newly diagnosed with breast cancer, one lady is dying of it [2, 3].

There are large number of algorithms for predicting and classifying the cancer. The current paper gives the prediction of breast cancer using machine learning with help of artificial neural networks in which feed-forward and back-propagation algorithms are applied. Our goal is to evaluate effectiveness and efficiency of neural networks for the prediction. The following questions are clarified in this study: How does Artificial Neural Networks exploits better effectiveness? Is algorithm efficient for the prediction? What accuracy does the algorithm provides?

II. Methodology Used

A. Neural Networks

A neural network is a machine that is designed to model the way in which the brain performs a particular task or function of interest; the network is usually implemented by using electronic components or is simulated in software on a digital computer.

B. Artificial Neural Networks

An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output.

C. Proposed Approach

In this research work, back propagation method was used in diagnosis of breast cancer. The method is classified in following stages. During primary stage, the network is trained by input data set with help of feed forward algorithm. The process is followed so that we obtain a best neural network. In the further stage, the breast cancer data is classified according to neural network model.

D. Back-Propagation

In the area of medical science, multi- layer neural network with back-propagation process are used mostly [5]. In Back-propagation methods, there are networks which travel from input neuron to an output neuron only in one direction. Back-propagation exist at least three layers of neurons as shown in Figure 1.

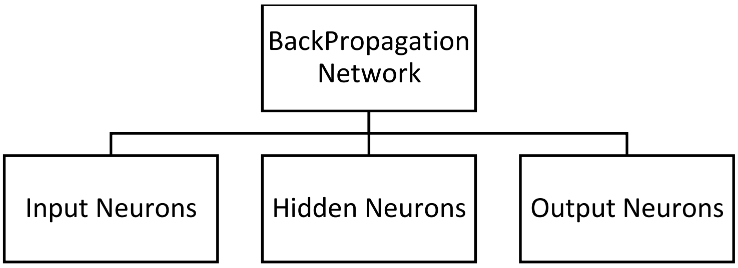

Figure1: Types of Neurons in Back-Propagation Network.

Neural network is basically a classified topology which constitute of neurons and connection based on weight. Each neuron in layer inside neural network are connected through synapses.

E. Neural Network Construction

For classifying the task a three layered back-propagation was implemented. The input layer consist of nine neurons .The hidden layer has 7 neurons. Hidden nodes can be varied among 4, 5, 6, 7, and 10. A series of trial has been done to achieve optimization of results. Accuracy of network training model was obtained best when hidden nodes were taken as 7. The output neurons has double nodes which represents diagnostic outcome (0, 1) for benign and (1, 0) for malignant.

Table1: Values of attributes used

Attributes	Value
Learning Rate	0.2
Number of Training Samples used	372
Number of Testing Samples used	100
Activation Functions	Sigmoid Function

Every set of attributes was inserted into the network on each iteration. This network was implemented on the Windows 10 Operating System using C compiler as the software for neural network training software.

III. Algorithm based on Back-propagation

In back-Propagation method, the error calculated at the output layer is send backward to the hidden layer from the output layer and then to the input layer. Each of the loop in back-propagation has two steps: feed-forward in which the information travels only in forward direction and the backward propagation to modify the weights by using the calculated error. The above mentioned steps are performed continuously until the ANN is satisfied with the optimum value within a value that is pre-specified. Back-propagation does the essential weight modification. [6]. The back-propagation has two parts: training of algorithm and constitutes its application.

Best neural network model is created using training algorithm. This training algorithm mainly comprises of following parts feed forward, error calculation and weight update.

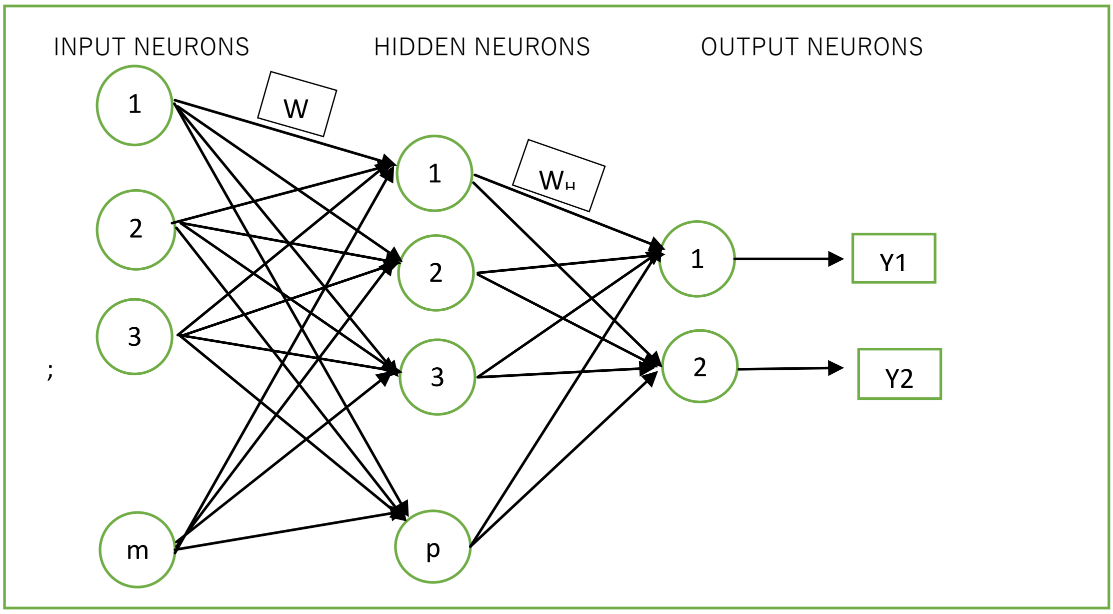

Figure 2: Pictorial representation of a three-layer artificial neural network with inputs neuron (sx_m), hidden neurons (W_{ih}) and output neurons (W_{ho}).

To start with an algorithm, weights are arbitrary created for the multilayer neural network. The process is then continuously proceeded several times till the calculated mean-square error (MSE) is less enough.

Procedure

1. Suppose the training example is E and the actual output are C.
2. Calculate the feed-forward propagation of E through the network (calculate the weighted sums of the network h_i and the activations for every neuron). This will be calculated for all the hidden layers and the output layers in the network.

   Hidden layer neurons (h_i, i= 1, 2, 3 …, p) total input weight,

$$h_i = W_{inp\ bias} + \sum_{i=1}^{m} u_i W_{ih}$$

Here, we have used Sigmoidal function as an activation function

$$f(x) = 1/(1 + e^{-x})$$

So, hidden node activation values are calculated as:

$$u_i = f(h_i) = 1/(1 + e^{-h_i})$$

Each output neurons (O_j, j= 1, 2, 3 …n) total the input weight and output node values are calculated as:

$$O_j = w_{out\ bias} + \sum_{j=1}^{n} h_i\, w_{h,o}$$

Thus, output node activation values are:

$$u_0 = f(o_j) = 1/(1 + e^{-o_j})$$

3. After the first pass, the error will be remarkable, but we will use back-propagation algorithm to adjust the weights so that it reduces the error between the wanted output and actual output. The error is calculated as follows:

$$Error = 0.5 * (O_{desired} - O_{actual})^2$$

For the output neuron $\delta_o = (C_i - u_0)u_0(1 - u_0)$

where $u_0(1 - u_0)$ is the sigmoidal derivative.

For all the Hidden Neurons

$$\delta_i = (\textstyle\sum_{m:m>i} w_{m,j})u_i(1 - u_i)$$

where m= All neurons connected to the hidden nodes, w= weight vector

4. Then the weights from the network are updated as:

Weights from hidden to output $w_{ij}^* = w_{i,j} + \rho\delta_0 u_i$

Weights from hidden to input $w_{i,j}^* = w_{i,j} + \rho\delta_i u_i$

Here ρ represents the learning rate.

The feed-forward algorithm calculates neurons activations and an actual value. The backward pass calculate the error. According to this, weights modify so that the error is reduced. The change that may take place for weights is reduced by learning rate.

Inputs are now been propagated to the hidden layer the final step is to feed the hidden layer. Then values are forwarded to the output layer to calculate the output of the network.

IV Experimental Results and Analysis

The experiment explained in this paper was programmed using C-language compiler. Different techniques are implemented using different libraries in other available machine languages but we preferred to apply using C-language as it helps to understand the depth of the subject. These algorithms are also useful in applying to other real world life problems.

A. Dataset based on Breast Cancer

The dataset used in this research has been taken from the UCI Machine Learning Repository, namely Wisconsin Breast Cancer (Diagnostic) dataset that was obtained by Dr. William H. Wolberg at the University of Wisconsin Madison Hospitals. 699 records have been used in this dataset. The dataset consists of nine features. These features are explained in Table 2 and are classified on basis of a period scale from 0.1–1.0. Here 1.0 represents the most abnormal state. Dataset that has been used has 699 cases (Malignant: 241, Benign: 458), 2 category (34.5% benign and 65.5% malignant) [4].

Table 2: Description of Characteristics of Wisconsin Breast Cancer Dataset.

Characteristic Number	Description of different Characteristics	Values
1	Thickness of Clump	0.1-1.0
2	Cell Size Uniformity	0.1-1.0
3	Cell Shape Uniformity	0.1-1.0
4	Marginal Adhesion	0.1-1.0
5	Size of Single Epithelial cell	0.1-1.0
6	Bare Nuclei	0.1-1.0
7	Bland Chromatin	0.1-1.0
8	Normal Nucleoli	0.1-1.0
9	Mitoses	0.1-1.0

B. Artificial Neural Network Training and Testing

The model was analysed and trained taking 372 instances from the dataset in which the benign and malignant data is (50%-50%) to make sure there is no biasing. These 372 instances where properly mined from the dataset so that there is greater accuracy and less error. About rest 30% of the data was tested and the network predicts the accuracy up to 98%.

The model was trained with the help of a back-propagation algorithm. This help in generating sufficiently less value for MSE.The algorithm described for trainingmodifies the weights to reduce the error from the network. After the network is trained, the weights are fixed. When the trained network is tested using some testing data values then it generates the output.The trained network gives the output which is either malignant or benign.

Table 3: Results of Prediction Accuracy obtained

Model No.	Model based on Hidden Neurons	Accuracy of prediction obtained (%)
1	No. Of Neurons =10	98.9247%
2	No. Of Neurons =7	99.4624%
3	No. Of Neurons =6	97.8495%
4	No. Of Neurons =5	98.6559%
5	No. Of Neurons =4	97.8495%

The simulation experiments were performed with 372 records and 9 attributes using Wisconsin breast cancer dataset. Table 3 shown above is describing the correction prediction percentage for every model that is tested on the trained network with 100 data as testing data. From the Table 3, Results clearly specifies that model 2 has obtained the maximum percentage of correction prediction, so we have achieve that the prime neural network model is model no 2 having Hidden Neurons equal to 7.

V. Conclusion

In this paper, a system of diagnosis for prediction of breast cancer was done perfectly using back-propagation algorithm. The dataset was trained and tested taking distinct neurons in hidden layer on which output was analysed. Back-propagation algorithm is used to test the model as outcome categorizer. From the experimental analysis, we obtained that model which has 7 hidden neurons has the highest precision as compared with the other models. The reason behind this is the mean square error. Lesser the mean square error more is the accuracy of the model and the error in this experiment is reducing towards 0.0001. The proposed feed-forward back-propagation algorithm can be considered a best classifiers as it predict the breast cancer disease with good accuracy.

References

[1] Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., ... & Bray, F. (2015). GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013.

[2] Bray, F., Ren, J. S., Masuyer, E., & Ferlay, J. (2013). Global estimates of cancer prevalence for 27 sites in the adult population in 2008. International journal of cancer, 132(5), 1133-1145.

[3] http://www.breastcancerindia.net/bc/statistics/stat_global.htm.

[4] Frank, A., & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive. ics. uci. edu/ml]. Irvine, CA: University of California. School of information and computer science, 213, 2-2.

[5] Burke, H. B., Goodman, P. H., Rosen, D. B., Henson, D. E., Weinstein, J. N., Harrell Jr, F. E., ... & Bostwick, D. G. (1997). Artificial neural networks improve the accuracy of cancer survival prediction. Cancer, 79(4), 857-862.

[6] Güler, I., & Übeyli, E. D. (2003). Detection of ophthalmic artery stenosis by least-mean squares backpropagation neural network. Computers in Biology and Medicine, 33(4), 333-343.