# BREAST CANCER DETECTION USING NEUTRAL NETWORKS

NEURAL NETWORK AND FUZZY SYSTEMS COURSEWORK-I

UoB# 14031187

## ABSTRACT:

In this modern era, medical science is still struggling in discovering breast cancer before it starts to spread. This is the main obstacle in recent times. So it was suggested to use automated systems. Different techniques were suggested, one of those suggested techniques was using artificial neural network. This is the technique in which a system can be trained by exposing this system to the training dataset. So afterwards when any case arrives, this system will be able to solve it efficiently. This whole process of training and using of this system is discussed in this report.

## INTRODUCTION:

Second leading cause of death of women is breast cancer [1]. Among every 500 women, one is effected by breast cancer in her life sooner or later [2]. The only possible solution to this problem is to diagnose breast cancer as soon as possible. It increases the chance of survival as nearly one third of women die due to breast cancer [3]. Many tools and systems were developed for doctors to decide whether patient has breast cancer or not. A neural network can be trained from dataset, and then for some unseen case it can give result in the form of some number for decision.

## BACKGROUND:

### OVERVIEW:

The basic model of a neuron is shown in below figure.

inputs    weights

1

$x_1$

$x_2$

$x_n$

$w_0$

$w_1$

$w_2$

$w_n$

weighted sum

$\Sigma$

step function
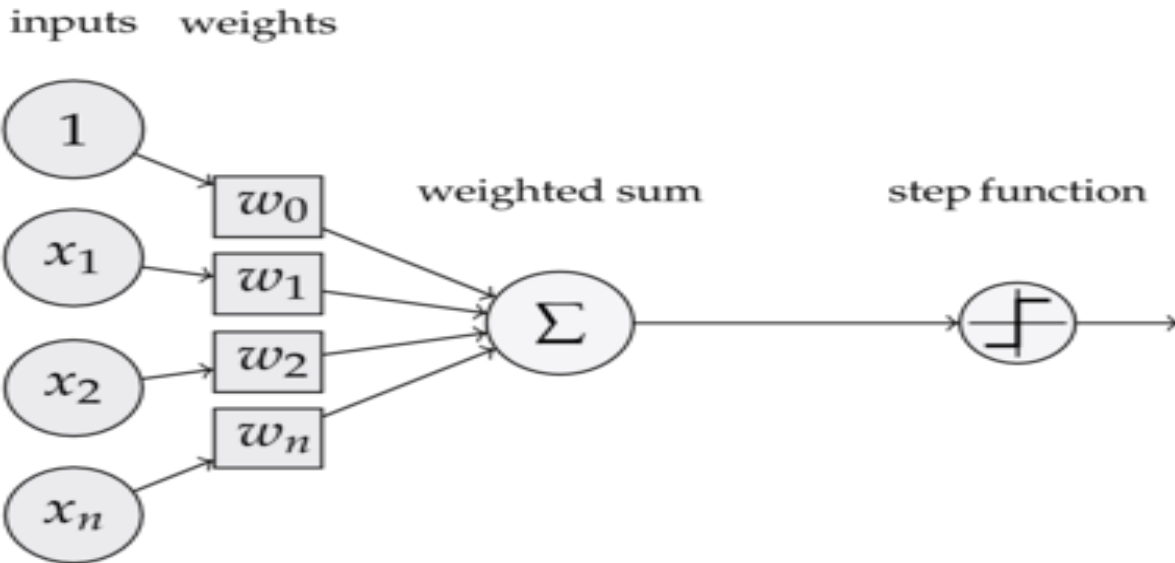
Input sets are applied to neuron and every input has some weight on it. The sum function add these input values after multiplying them with their associated weights. Primarily, the weights are randomly picked. Them the sum function will pass the output to activation function from where the actual output will be generated. There are several activation functions that produces more useful output. As, the binary threshold generates the output as 1 or 0. In the same way, other activation functions also works according to their functionality like linear function and sigmoid function [4].

Artificial neural network is a combination of neurons. Basically it consists of three layers, called the input layer, hidden layer and output layer. Every layer has some nodes called neurons which are bound to each other. The flow of data is shown by the lines between nodes from input to output nodes that is from left to right. There are two possible classes of ANN based on their pattern: feed forward networks and recurrent networks. In feed forward networks, neurons are arranged in layers first deals with input, second deals with output while the middle layer called hidden layer, has no connection with outside. Moreover, there is no connection among neurons in the same layer and there is no involvement of previous state of neurons. It is easy to implement. On the other hand, in recurrent networks, state of neurons depends upon inputs as well as previous state of neurons. This algorithms is used to improve the performance [5].
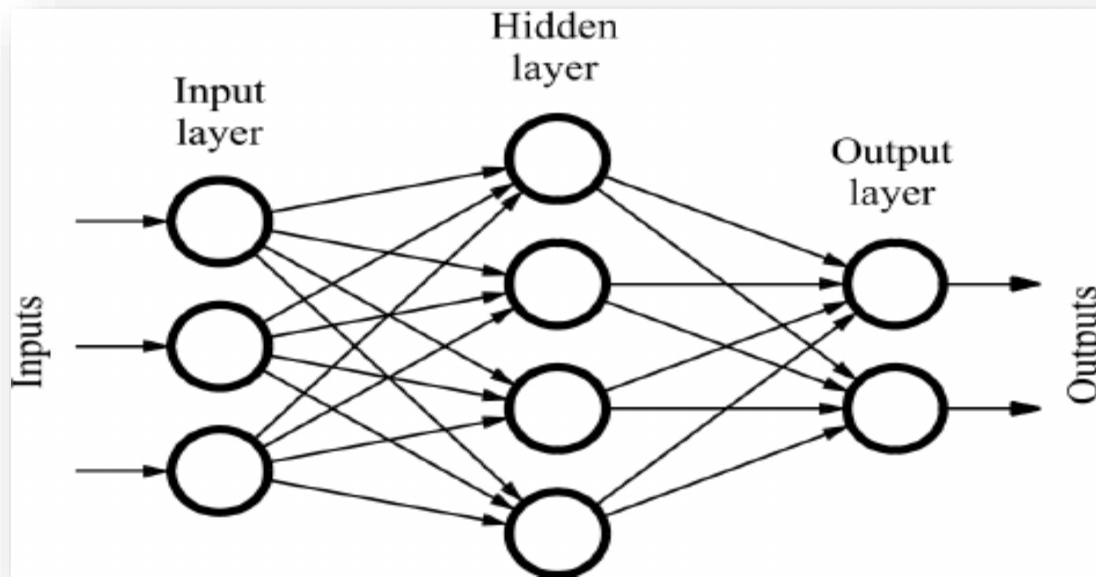
The very important part of neural network are hidden layers. It changes the inputs so that output layer can use it easily. One cannot get the number of layer or neurons in it theoretically. Moreover, there should be same number of inputs and outputs to the neurons in the input and output layers respectively.

## NEURAL NETWORK TRAINING:

Artificial neural networks inspires from biological nervous system and the most interesting quality of human brain is that it learns, it learns from surroundings and then make decisions according to them. In the same way, artificial neural networks are being trained by some training dataset and then they make quick and intelligent judgment on new cases [4]. The simplified neural network can be trained by adjusting the weights, this will reduces the errors and helps the network to fulfil the accuracy goal as a result the performance of network also improve. Accuracy also depends upon many things like hidden layer, learning rate and activation function which will be discussed later in experiment section.

## PREVIOUS WORK:

Artificial neural network can detect breast cancer in a better way than any human as understanding of data pattern is much better in neural network. Likewise, it is capable of fast identification as it is Infallible of human errors and emotions. In 2012, a 17 years old boy, *Brittany Wenger*, global neural network based in the cloud that assist the physicians to diagnosed the cancer. It detect malignant breast tumor with the accuracy of 99.11% [6]. Another work is done

by *Kala* and his fellows in 2011 to diagnose breast cancer. Evolutionary neural network architecture is proposed and its performance in training data is 96.28% and on 95.78% testing data [8]. In 2010, *Janghel* implemented four neural network model, competitive learning, back-propagation, radical basis functions and learning vector quantization then perform analysis on them. The models on the basis of their performance on testing data are LVQ, CL, MLP and RBF.

## *MAIN PART:*

I will discuss the phases to solve this problem we clean the provided data and make it able to deal with neural network. Then I will discuss the whole process from creating the neural network to finding the accuracy rate.

In this assessment, the problem I was asked to solve is to diagnose breast cancer by separating benign and malignant tumor. For this the dataset of 699 breast cancer's patients was provided which is obtained from *University of Wisconsin Hospitals*. The data contains patient's useful data for processing and as well the actual output. The data was in text file so first I make the data able to deal with neural network. I converted the text file into an array in matlab file and it was mentioned that the first attribute is patient's Id and last attribute is output which is 2 or 4, 2 for benign and 4 for malignant. So these two attribute were detached from dataset and the output attribute was saved for further analysis. There are some "?" in data due to which I have omitted this data and total remaining dataset is 683.

As we know that classification problems can be explained by static neural network and as well as dynamic. So, I created feed-forward back-propagation network by '*newff*' command. Other network's parameter are: epochs 200 and goal 0.01. The weights will be randomly selected by the networks. To determine the accuracy, I created accuracy function that take actual result and network's result as input and return the percentage accuracy.

## *EXPERIMENTAL RESULTS AND ANALYSIS:*

The accuracy of neural network varies by changing some values like number of hidden layers and neurons in them. Moreover, the division of also affects the accuracy. I will perform different experiments to find the best parameter's value to get maximum accuracy.

### Experiment I – Data Distribution Effect:

There are two types of data, one is for neural network training and other is for testing. The division of these data can vary the accuracy of our neural network.

Hypothesis:

By increasing training data, the accuracy will be increased.

PARAMETERS:

- Goal = 0.01
- Activation function = tansig, tansig
- Hidden layer = 20
- Validation checks = 10
- Default learning function

RESULTS:

| Training data | Testing data | Accuracy |
|---|---|---|
| 95% | 5% | 100% |
| 80% | 20% | 98.6% |
| 70% | 30% | 98.5% |
| 60% | 40% | 97.4% |
| 50% | 50% | 97.3% |
| 30% | 70% | 97% |
| 10% | 90% | 93.6% |

From above result it is clear that by increasing training data accuracy will be increased.

## Experiment II –Number of Neuron Effect:

The number of neuron in a single hidden layer affects the accuracy. More neurons are in layer the more deep learning will occur hence accurate will be increased.

## Hypothesis:

By increasing neurons, accuracy will be increased.

## PARAMETERS:

- Goal = 0.01
- Activation function = tansig, tansig
- Hidden layer = 20
- Validation checks = 10
- Default learning function
- Data distribution: training = 50%, testing = 50%

RESULTS:

| No. of neurons | Accuracy |
|---|---|
| 3 | 93.2% |
| 5 | 94.2% |
| 8 | 95% |
| 12 | 97.8% |
| 20 | 98.8% |
| 30 | 98.5% |
| 50 | 98.5% |

From the above result, it is clear that accuracy is increased by increasing number of neurons to some instance but frequency is much less.

## Experiment III – Change in goal Effect:

The goal affects the accuracy. The higher the goal, more chance on error due to which accuracy will be decreased.

## Hypothesis:

By increasing goal, accuracy will be decreased.

## PARAMETERS:

- Goal = 0.01
- Activation function = tansig, tansig
- Hidden layer = 20
- Validation checks = 10
- Default learning function
- Data distribution: training = 50%, testing = 50%

## RESULTS:

| Goal | Accuracy |
|---|---|
| 0.001 | 95% |
| 0.01 | 94.7% |
| 0.02 | 93.6% |

From the above result, it is clear that accuracy is decreased by increasing goal.

## CONCLUSION:

The purpose of diagnosing breast cancer by neural network is to get the more accurate result as no human can see the data pattern as compared to neural network. In this assignment I also generate a neural network that differentiate malignant and benign tumor with maximum accuracy. By experimental analysis I can say that accuracy rate is directly proportional to the training data. With 95% of training data along 5% testing data the accuracy was 100%, however the 60-70% training data is fair enough for neural network along 40-30% testing data respectively. Likewise accuracy also increases with the increase of no. of neurons but at some instance. But it's not the all, there are some other factors that also changes the accuracy.

# *BIBLIOGRAPHY*

1. Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E., & Thun, M. J. (2003). Cancerstatistics. Cancer Journal for Clinicians, 2003(53), 5–26

2. J.R. González, V. Moreno, E. Fernández, A. Izquierdo, J. Borrás, R. Gispert,Probabilidad de desarrollar y morir por cáncer en Cataluña en el periodo 1998–2001, Med Clin (Barc), 124 (2005), pp. 411–414

3. Scheidhauer, K., Walter, C., & Seemann, M. D. (2004). FDG PET and other imagingmodalities in the primary diagnosis of suspicious breast lesions. EuropeanJournal of Nuclear Medicine and Molecular Imaging, 31(Suppl. 1), 70–79.

4. Lee Jacobson. (2013). Introduction to Artificial Neural Networks - Part 1. [ONLINE] Available at: http://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-1/7. [Accessed 9 December 2016].

5. Saritas, I. J Med Syst. (2012). Prediction of Breast Cancer Using Artificial Neural Networks, Journal of Medical Systems, 36(5):2901–2907, 2012.

6. Saritas, I. J Med Syst. (2012). Prediction of Breast Cancer Using Artificial Neural Networks, Journal of Medical Systems, 36(5):2901–2907, 2012.

# APPENDIX:

```matlab
data = importdata('breast-cancer-wisconsin.data');
list = [68,204,342,410,478,546,648];
input = data(:,2:10);
output = data(:,11);
traindatafirst = input(1:list(3)/2,:);
traindatalast = input(size(input)-list(3)/2+1:size(input),:);
targetdatafirst = output(1:list(3)/2,:);
targetdatalast = output(size(output)-list(3)/2+1:size(output),:);
testinputData = input(list(3)/2+1:size(input)-list(3)/2,:);
testouputData = output(list(3)/2+1:size(output)-list(3)/2,:);
concattraininginput = cat (1,traindatafirst,traindatalast);
concattrainingoutput = cat(1,targetdatafirst,targetdatalast);
net = newff(concattraininginput',concattrainingoutput',20, {'tansig'
'tansig'}, 'trainr', 'learngd', 'mse');
net.trainParam.goal = 0.1;
net.trainParam.epochs = 200;
net.trainParam.max_fail = 6;
net = train(net, concattraininginput',concattrainingoutput');
malik = net(testinputData');
malikS = malik';
counter = 0;
for i=1:size(malikS)
      if (malikS(i)<=3)
            malikS(i)=2;
      else
            malikS(i)=4;
      end
      if(malikS(i)== testouputData(i))
            counter=counter+1;
      end
end
      percentage=counter/size(testouputData,1)*100;
      disp(percentage);
```