BREAST CANCER CLASSIFICATION USING NEURAL NETWORKS

Abstract

Artificial intelligence is an evolving aspect of computer science. One thing that make humans intelligent is our ability to learn from experience and from people around us. To make software self-adoptive, we use autonomous systems so that it become independent from human dealing. Breast cancer is one of the hot topics of study that need this kind of system/software. Lot of ideas have been proposed for the detection of this disease. Artificial Neural Networks can be used for this purpose by training it with a dataset in such a way that it can be able to classify when a new case arrives. This report is about providing a solution that how Neural Networks can help us in "Breast Cancer Data Classification".

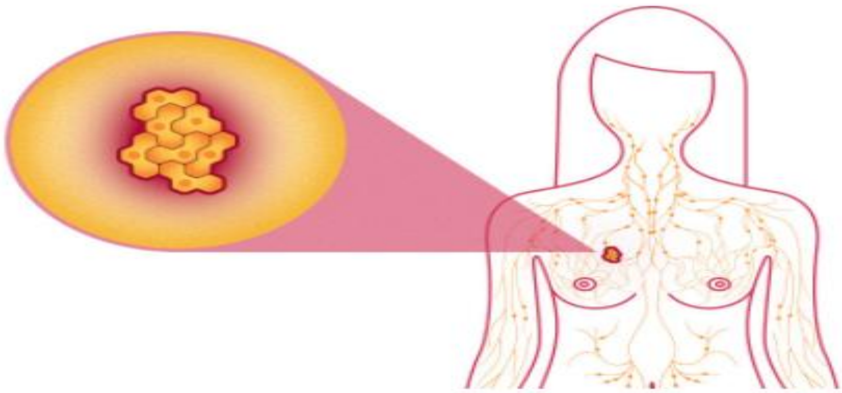
Nisar Ali 14031279

nisar2014@namal.edu.pk

Table of Contents

1. What is Breast Cancer

Breast cancer is a type of disease where malignant (cancer) cells form in the tissues of the breast. It starts when breast cells begin to grow out of control. A tumor appeares due to these cells which can easily be seen on an x-ray. Breast cancer exists almost entirely in women, but there is a chance that men can get breast cancer, too [1].



2. Introduction

Among women, the most commonly diagnosed cancer is of the breast. In United States, one in eight women will be diagnosed with this disease in her lifetime. This disease is the 2nd leading cause of cancer deaths in women. Each year over 252,500 women in United States will be diagnosed with breast cancer and more than 40,000 will die due to this disease. This disease is rare in men but it is estimated that over 2,500 men will be diagnosed with this disease and 500 will die every year. On average, every 2 minutes a woman is diagnosed with this fatal disease and a woman will die every 14 minutes. Over 3.3 million survivors are alive in United States with breast cancer [2].

A lot of work has been done to stop this disease from causing harm. It takes time to diagnose breast cancer due to which it spreads and the chances of death increases. It is very important to diagnose such a fatal disease quickly and for this there should be a computerized system which will help to diagnose breast cancer faster. For this many techniques have been proposed and one of this is Artificial Neural Networks. This method have been used widely to classify breast cancer to help the community.

3. Background

As I have discussed above that there should be a computerized system to diagnose breast cancer. Neural networks have played a very important role for the past few years in this field. Now I am going to discuss that what is neural network and how it works.

The artificial neural network/model is said to be a system processing input signals into a single output. Artificial neuron is the very first step of an artificial neural network. Artificial neurons have n inputs and one output. The information about input is always numeric and forms the vector of input values. For every neuron there is a numerical factor called weight. Input values of weight are different and generated automatically in the learning process. The neural operation consists of two phases, the first one is defined as an aggregation of input value which is calculated by multiplying the input values and the weights. In the second phase the aggregated sum of values becomes an argument in the activation function. At the very start of the process, the input variables get initially randomly assigned weights. The proper values of weighting factors can be determined in the learning process.

Activation function produces an output within a certain threshold. There are several types of activation function. Among the number of activation functions, the most commonly used are: Linear activation function, sigmoid activation function, Gaussian activation function, hyperbolic tangent activation function [3]. Figure 1 shows the basic model of artificial neuron.

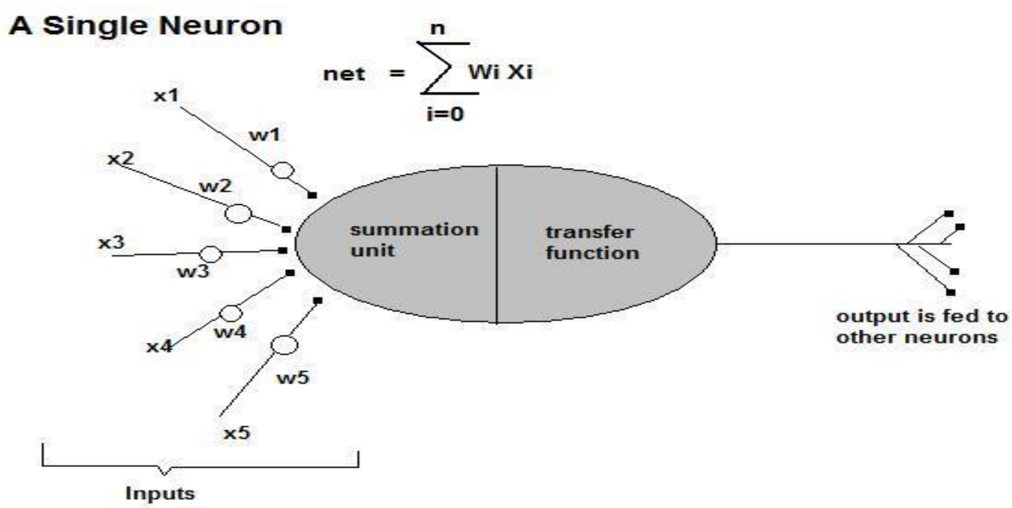


Figure 1: Model of a Single Neuron

4. Artificial Neural Network

Artificial neural network has provided an exciting alternative method for solving a variety of problems in different fields of science and engineering. An artificial neural network is defined as the interconnection pattern between the different layers of neurons. Each connection between neurons has a weight [8]. Each neuron is represented by a node and these nodes are connected through edges which makes the output of one neuron, the input of other. Two main types of neural networks are given below:

- ✓ Feed Forward network
- ✓ Feedback network

4.1. Feed Forward Network

Feed forward network consists of perceptron which are organized in the form of layers. A feed forward network is a simple type of neural network which consists of an input layer, an output layer and one or more hidden layers. Each perceptron in a single layer has no connection with each other. Furthermore each perceptron in a layer is connected to all perceptron's on the next layer. The main advantage of this network is that it learns to evaluate and recognize input patterns [4]. In this network the information moves only in one direction i.e. forward, from the input nodes data goes through hidden nodes (if any) and then to the output nodes. Figure 2 shows the basic model of feed forward network.

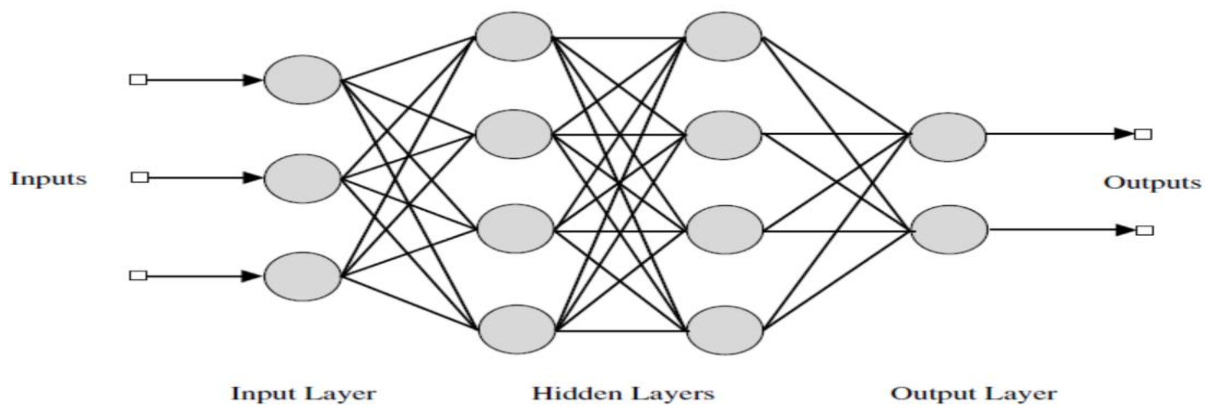


Figure 2: Basic Model of Feed Forward Network

4.2. Feedback Network

In this type of network, the output goes back into the network to achieve the best result. Feedback networks are dynamic. Signals are traveling in both directions by introducing loops in this network. The network remains at the equilibrium point until the input changes. Feedback networks are used by the internal systems for the correctness of errors [4]. Figure 3 shows the basic model of feedback network.

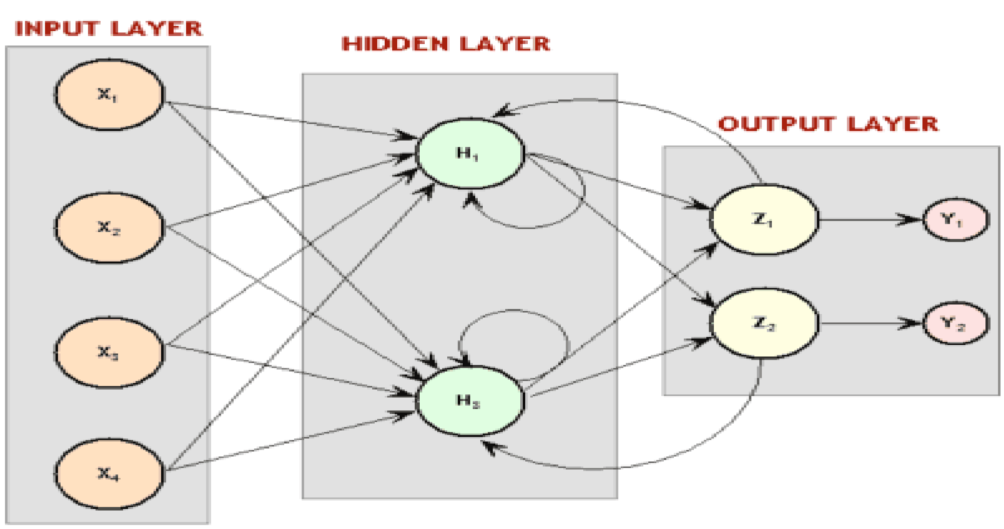


Figure 3: Basic Model of Feedback Network

5. Artificial Neural Network Training

Artificial neural network is a computing system inspired by human brain. A human brain learns from its surroundings and makes decision using knowledge. An artificial neural network can be developed just like a human brain by training it with proper dataset so that it can be able to solve unseen problems related to the specific dataset.

A neural network can be trained by different methods. Feed forward network can be trained by using Conjugate Gradient and Back Propagation (BP) method, the most common learning algorithms. On the other hand feedback network can be trained by using real-time recurrent learning algorithm [9].

6. Related Work

In the last decade, a lot of work has been done for the detection of breast cancer. Today, a number of screening techniques are used to detect breast cancer like positron emission tomography (PET), CT scan, X-Ray, ultra sound, mammogram etc. These techniques have their own advantages. Mammogram is the most reliable and popular technique. But, this technique has some serious limitations. About 30% of the breast grazes couldn't be spotted in mammogram during screening. This thing led the researchers towards developing an automated computational system for breast cancer diagnosis [10].

Artificial neural networks are used for the diagnosis of breast cancer since the past few years and their accuracy rate is very significant as compared to other non-computational diagnostics. For classification Support Vector Machine (SVM), Self-Organizing Map (SOM), Probabilistic Neural Network, General Regression Neural Network, Radial Basis Network (RBN) and Multi-Layer Perception (MLP) have also been applied on the said task. The obtained results show that General Regression Neural Network has been the most accurate rate in identifying the nature of the input as compared to others. Its accuracy rate was 98.8% respectively [10].

7. Main Part

This part contains the description of methodology developed for the classification of breast cancer.

7.1. Dataset Management

We have a dataset of breast cancer which was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. Furthermore, the dataset was converted from a text file to ".m file" (Matlab file). As it was mentioned in the dataset documentation that the first column shows the id number of patients, so that column was deleted from the dataset. There were some missing values in the dataset which were deleted (with rows) from dataset. The last column in the dataset consisted of 2's (benign) and 4's (malignant) and was replaced by 0 (benign) and 1 (malignant). After doing this, the data was divided into two parts i.e. training data and testing data.

7.2. Solution to the Problem

For the classification of breast cancer data, I have written the code in Matlab and the neural network was created for training and testing purpose. The neural network used for this problem is feed forward back propagation network. In neural network different combinations were used for the training and testing data like different number of hidden layers and neurons, learning functions, training functions, activation functions etc.

8. Experimental Results and Analysis

In this section, each experimental result will be analyzed by checking the performance of neural network through accuracy.

8.1. Data Distribution Effect

The dataset given to us will be divided into different ratios for this particular experiment. The activation, learning, training and performance functions are “tansig, purelin”, “learngd”, “trainr” and “mse” respectively with four hidden layers. Each hidden layer has different number of neurons.

The expected result of this experiment is that “as the ratio of training data decreases the accuracy will also decreases”. Table 1 will show the accuracy according to the training and testing data:

Training Data	Testing Data	Accuracy
90%	10%	99.6%
80%	20%	99.55%
70%	30%	99.51%
60%	40%	98.9%
50%	50%	97.36%
40%	60%	97.32%
30%	70%	97.2%
20%	80%	96.5%
10%	90%	95.8%

Table 1: Data Distribution Effect

8.1.1. Analysis

It can be seen from the above experiment that the hypothesis was true as the performance of neural network (accuracy) decreased by decreasing training data. It is because of that the neural network has a property, the more the training data is increased the better accuracy gets.

8.2. Activation Functions Effect

In this experiment different combinations of activation functions will be used to check the performance of neural network. Other functions and hidden layers are the same as in first experiment. In this experiment I have set the training data to 70% and testing data to 30%, which is a standard distribution. Table 2 will show the accuracy according to the activation functions.

Training Data	Testing Data	Activation Functions	Accuracy
70%	30%	tansig, tansig	99.01%
70%	30%	tansig, purelin	99.8%
70%	30%	purelin, purelin	98.5%
70%	30%	purelin, tansig	99.02%
70%	30%	logsig, logsig	99.5%
70%	30%	logsig, purelin	99.01%
70%	30%	compet, compet	79.3%
70%	30%	compet, tansig	76.8%
70%	30%	hardlim, hardlim	86.7%
70%	30%	hardlim, hardlims	94.9%

70%	30%	hardlims, hardlims	95.6%
70%	30%	hardlims, hardlim	89.2%
70%	30%	poslin, poslin	99.7%
70%	30%	poslin, hardlims	98.03%
70%	30%	satlins, satlins	99.5%
70%	30%	satlins, poslin	99.4%
70%	30%	tribas, tribas	98.5%
70%	30%	tribas, satlins	99.6%

Table 2: Activation Functions Effect

8.2.1. Analysis

It can be concluded from the above experiment that the neural network will give its best performance when the activation functions are "tansig, purelin". It is because of that the tansig and purelin functions resulted in the maximum and the minimum computational cost which indicated the simplicity and complexity of these functions during the processing [5]. So here our hypothesis is validated that "by changing activation functions we can get better accuracy".

8.3. Effect of Increasing Hidden Layers

Different number of hidden layers with different number of neurons will be checked in this experiment to check the best performance of neural network on data. Again in this experiment the training and testing data will be 70% and 30% respectively. The activation, training and learning functions will be "tansig, purelin", "trainr" and "learngd".

The expected result of this experiment is that "as the number of hidden layers increases the accuracy will also increases". Table 3 will show the accuracy regarding to hidden layers.

Training Data	Testing Data	No. of Hidden Layers	Accuracy
70%	30%	1	98.21%
70%	30%	2	98.52%
70%	30%	3	98.9%
70%	30%	4	99.01%
70%	30%	5	99.31%
70%	30%	6	99.53%
70%	30%	7	99.70%
70%	30%	8	99.8%

Table 3: Effect of Increasing Hidden Layers

8.3.1. Analysis

If we look into table 3, we have seen that the accuracy increases as the number of hidden layers increases. So our hypothesis is true. It is because more hidden layers means, more non-linearity applied to data before the last layer. In other words we can say that by adding more layers to a neural network adds dimensionality to solve more complex problems [6].

8.4. Training functions Effect

Different combinations of training functions will be used to check the performance of neural network in this experiment. Other functions and training, testing data are the same as in above experiment. The hidden layers are four with different number of neurons. Table 4 will show the accuracy according to the training functions.

Training Data	Testing Data	Training Functions	Accuracy
70%	30%	trainbfy	99.01%
70%	30%	traincgb	99.51%
70%	30%	traincgf	99.02%
70%	30%	traincgp	99.02%
70%	30%	trainlm	98.03%
70%	30%	trainoss	99.02%
70%	30%	trainrp	99.51%
70%	30%	trainscg	99.01%
70%	30%	traingdm	99.5%
70%	30%	traingd	99.01%

Table 4: Training Functions Effect

8.4.1. Analysis

In table 4, training functions which are listed are best among all training functions. This is because their testing mean time is very low as compared to others. The "trainrp" function has low testing mean time (0.0229) and that is why its accuracy is better than others [11]. So our hypothesis is true here that "by changing different training functions we can get better results".

8.5. Malignant and Benign Rows Effect

Our training and testing data is same as in other experiments. But, in this experiment the ratio of malignant and benign rows will change. All other functions and hidden layers are same for this experiment. The hypothesis is that when the ratio of benign rows is more than malignant rows then our neural network will give us better accuracy. Table 5 will show all the effects.

Training Data	Testing Data	Malignant Rows	Benign Rows	Accuracy
70%	30%	20%	80%	95.7%
70%	30%	30%	70%	97.5%
70%	30%	40%	60%	98.5%
70%	30%	50%	50%	96.5%
70%	30%	60%	40%	97.8%

Table 5: Malignant and Benign Rows Effect

8.5.1. Analysis

We can conclude from the experiment that the performance of neural network is better when the ratio of benign rows are more than malignant rows. When the ratio of benign rows is 60% and malignant rows is 40% then the accuracy is better. It can also be seen in the table that the ratio of benign rows is way more than the malignant rows, the performance of system is low due to most errors will occur in the case of malignant rows.

8.6. Learning Functions Effect

In this experiment, I have used different learning functions to check the best performance of neural network. All other settings are same as for other experiments. Table 6 will show the performance of neural network.

Training Data	Testing Data	Learning Function	Accuracy
70%	30%	Learngd	99.01%
70%	30%	Learncon	99.02%
70%	30%	Learngdm	98.5%
70%	30%	Learnnp	99.01%
70%	30%	Learnpn	98.5%
70%	30%	Learnwh	99.5%

Table 6: Learning Functions Effect

8.6.1. Analysis

It is concluded from the experiment that by changing and testing different learning functions we can get the best performance of neural network. As "learnwh" learning function has the best accuracy among others and it is because of the learnwh calculates the weight change dW for a given neuron from the neurons input and error, and the weight learning rate LR, according to the Widrow-Hoff learning rule. It is also known as delta or least mean squared (LMS) rule [12].

dW = lr*e*pn'

8.7. Effect of Increasing Number of Neurons

The number of neurons in a single layer will be increased in this experiment just to check the performance of neural network. All other functions and data are set to default as for other experiments.

In this experiment our hypothesis is that "by increasing the number of neurons in a single layer, we will witness increase in the accuracy of neural network". Table 7 will help us to find our hypothesis true.

Training Data	Testing Data	Number of Neurons	Accuracy
70%	30%	1	70.4%
70%	30%	4	76.2%
70%	30%	8	98.5%
70%	30%	12	99%
70%	30%	16	99.5%
70%	30%	20	99.8%

Table 7: Effect of Increasing Number of Neurons

8.7.1. Analysis

When the number of neurons in a single layer are 1 and 4, the accuracy is very low and there are chances of errors. The reason for this is that the number of neurons that take the data from input vectors are less, and that is why they failed to transfer correct information to the next layers. The results are acceptable with 12 neurons in hidden layer and by increasing the number of neurons further, no further improvement was observed [13]. So again our hypothesis is true here.

8.8. Performance functions Effect

We have to check different performance functions in this experiment. Table 8 will show all the results.

Training Data	Testing Data	Performance function	Accuracy
70%	30%	mae	99.51%
70%	30%	mse	99.5%
70%	30%	msereg	97.5%

70%	30%	ssc	87.7%

Table 8: Performance Functions Effect

8.8.1. Analysis

From the above experiment we can say that by changing/testing different performance functions, we can get better performance. The “msereg” function measures network performance as the weight sum of two factors: the mean squared error and the mean squared weight and bias values. That is why its accuracy is better than other performance functions [7].

9. Conclusion

The report sheds light upon different ways in which a better result can be found for the classification of the breast cancer. Neural network depends on data given to it and on the basis of data, training and testing data will be separated. In this report the training data is distributed to 70% and testing data to 30%. On this distribution of data the neural network is able to give 99% accuracy. Neural networks depend on activation, training, and learning and performance functions. Bearing all this in mind, our neural network will be able to give us a better performance.

10. Bibliography

1. Cancer.org. (2017). *How Common Is Breast Cancer?*. [online] Available at: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html [Accessed 3 Dec. 2017].
2. www.nationalbreastcancer.org. (2017). *Breast Cancer Facts :: The National Breast Cancer Foundation*. [online] Available at: http://www.nationalbreastcancer.org/breast-cancer-facts [Accessed 3 Dec. 2017].
3. Siderska, J. (n.d.). APPLICATION OF NEURAL NETWORKS FOR SOCIAL CAPITAL ANALYSIS.
4. Anon, (2017). [online] Available at: https://www.quora.com/What-are-the-various-types-of-neural-networks [Accessed 3 Dec. 2017].
5. Google Books. (2017). *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications*. [online] Available at: https://books.google.com.pk/books?id=FNC5BQAAQBAJ&pg=PA291&lpg=PA291&dq=why+tansig+and+purelin+activation+function+accuracy+is+better+than+others&source=bl&ots=0477oSqjRx&sig=mu4OtCeukCvv-Wqv69rGWW7-vJ0&hl=en&sa=X&ved=0ahUKEwiqntHW7-3XAhWKPY8KHU2PA6EQ6AEIPTAF#v=onepage&q=why%20tansig%20and%20purelin%20activation%20function%20accuracy%20is%20better%20than%20others&f=false [Accessed 3 Dec. 2017].
6. Anon, (2017). [online] Available at: https://www.quora.com/Why-do-neural-networks-with-more-layers-perform-better-than-a-single-layer-MLP-with-a-number-of-neurons-that-leads-to-the-same-number-of-parameters [Accessed 3 Dec. 2017].
7. Cens.ioc.ee. (2017). *msereg (Neural Network Toolbox)*. [online] Available at: http://cens.ioc.ee/local/man/matlab/toolbox/nnet/msereg.html [Accessed 3 Dec. 2017].
8. Samatin Njikam, A. N., & Zhao, H. (2016). A novel activation function for multilayer feed-forward neural networks. *Applied Intelligence*, *45*(1), 75–82. https://doi.org/10.1007/s10489-015-0744-0
9. Jiang, M., & Zhang, B. O. (2003). Fast Learning Algorithms for Feedforward Neural Networks, 37–54.
10. Pandey, B., Jain, T., Kothari, V., & Grover, T. (2012). Evolutionary Modular Neural Network Approach for Breast Cancer Diagnosis, *9*(1), 219–225.
11. Vacic, V. (2005). Summary of the training functions in Matlab's NN toolbox. *Matlab*.
12. Demuth, H. (2006). Neural Networks. *Mathworks Inc*, *19*(1), 1–7. https://doi.org/10.1016/j.neunet.2005.10.002
13. Jadhav, U., & Shetty, A. (2016). Effect of varying neurons in the hidden layer of neural network for simple character recognition. *International Journal on Recent and Innovation Trends in Computing and Communication*, *4*(6), 4–7. Retrieved from http://www.ijritcc.org/download/browse/Volume_4_Issues/June_16_Volume_4_Issue_6/1466658910_23-06-2016.pdf