

Introduction

Breast cancer is one of the main causes of death among women and the most frequently diagnosed non-skin cancer in women.¹ The cells of the body multiply uncontrollably in cancer, wild growth of the cells causes cancer. Mostly this disease occurs during the age of 40 to 50. In the beginning breast cancer starts from inner lining of milk ducts, from here it can be increased to the other parts as well.

In most cases size of the tumor tells about the stage of the cancer. There are different methods of staging the breast cancer, one of them is from stage 0 to 4, it can be broken into further stages as well. There are different types of breast cancer like Ductal carcinoma and Lobular carcinoma. Ductal carcinoma is the most common and it starts in the milk duct. Lobular carcinoma begins in the lobules.

Sometimes infected cells break out into the lobes which is the most dangerous aspect of the disease, because it can increase the chances of spreading the disease in other parts of the body. Men can also be affected by the breast cancer but chances are very less than in the women.²

In the beginning there was a method of mammography to detect the breast cancer but that was not so efficient and life saving rate was very low. Image segmentation was also used earlier to detect the bright spots of the disease. Artificial Neural Networks are computer systems that have been developed for the purpose of automatically performing skills such as generating new information through learning and forming and discovering new information, which are characteristics of the brain.³ In artificial neural networks there is some input data, hidden layers and the output data. Neural networks work under some restrictions and follow some boundaries. Training and testing is the prime purpose of the neural networks.

Coming on to the given problem, we have breast cancer data which was collected from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg and we have Neural Networks from Artificial Intelligence domain. We have to come up with a trained Neural Network which gives us a better/accurate classification rate.

Background

Breast cancer can either be malignant or benign and a potent killer, if left undiagnosed. Lots of hard work, energy and effort are expended to find better diagnostic techniques and cures for the deadly disease since timely detection can ensure effective treatment and better chances of survival. Naturally in the digital age, computer systems are employed to aid the cause by the use of several methods.

Neural networks have proven an efficient technique in the detection and classification of breast cancer, which mimics the abstraction of a working human brain. The smallest unit of the network is an artificial neuron, each with its distinctive weight organised abundantly in varying layers together to converge into a single numerical output. The algorithm is fed training data to improve its accuracy and yield better results.

The computational power of a neural network lies in the pathway of artificial neurons that bind together to produce a single output. As the training data increases, the prediction errors are minimised since every time a correct output is given, the specific path of neurons is strengthened to give us reinforcement learning, just like humans learn from experience.

Main Part

This section contains the description of the methodology, justification for design decisions and the system architecture.

Data given for the problem is not equally distributed. First of all sorting is done to the given data and save this sorted data in a some text file. In the given amount of data first nine columns are only the symptoms and the last column tells the story whether by using these symptoms a patient has cancer or not. I made the small chunks of the data for training my neural network. Making the chunks of the given information is the prime purpose and in those chunks my assumption is if I will take equally distributed values of given data of the those patients who has cancer and the patients who do not have the cancer for training my neural network, it will directly help my algorithm to train more efficiently. I took the small amount of data for training and test it for all of the given remaining information for the problem. I also used the algorithm in such a way that it takes values of those data of those patients who has the cancer from start of the sorted and also took the same amount of data of those patients who do not suffering from the cancer from end of sorted list.

While training the neural network different schemes was adopted like changing the goals and epochs etc. for training neural network and making it more efficient in results hidden layers also play an vital role. Results of the trained neural network was not in integers and most of the times it is in the form of floating values. In given data all of resultant values are either 2 or 4 form so conversion of all resultant trained floating values changed into integers by applying a simple loop.

Accuracy of the data also checked by applying a formula which uses the values of counter size and multiplication by 100. Accuracy directly affected by changing the iterations, goal and hidden layers. If the network reached to the goal before completing the iterations it will stop training and it may occur that all the iterations took place and goal is not achieved by the network. Hidden layers played an important role in this whole scenario.

Experimental Results and Analysis

Methodology explained in the above section is used to perform different experiments which gave variety of results. First of the all neural network trained on some random data and values and results noted down as it is, but after that few hypothesis were made and on the basis of those hypothesis some results is driven. Those hypothesis and their analysis are discussed below

Hypothesis 1

If the training data is increased than accuracy will increase with the same ratio.

Analysis

Different experiments performed and in each experiment amount of training data is increased in result of which accuracy increased. Hypothesis made for this was very correct according to the different experiments performed by neural network. While increasing the data set all of the other constraints remained same. Accuracy decreased with the same ratio as well while decrease in training data happened.

Hypothesis 2

If the epochs increased while training the neural network accuracy will give increment in its values.

Analysis

There is no significant change in the accuracy while changing the epochs with the small amount but there is change in the percentage of accuracy while a big and vital change made in the epochs. In most of the experiments values did not reached to the complete iterations because before that values met to validation check or goal.

Hypothesis 3

If the value of a goal increased there will be a fall down in accuracy of the neural network.

Analysis

Goal and the accuracy of neural network are inversely proportional in various experiments driven by network. Increment in the values of the goal gave less amount of accuracy to the system while other all subjects remained same. There is a proper decrease in accuracy noted down while goal is increased to some high values.

Hypothesis 4

If the hidden layers increased there will be noticeable increase in accuracy.

Analysis

Hidden layers play vital role in neural networks. They increase non-linear functionality of the network while increase in the hidden layer took place. On each hidden layers different weights are accountable for performing the different functions.

In the beginning of experiments there is only one hidden layer, which changed from one to four hidden layers while all of other values of constraints remain same as those were in only one hidden layer. Increase in hidden layer directly hit the accuracy value and there is significant increment in the accuracy value of the neural network.

Hypothesis 5

If the validation check decreased there will be decrease in accuracy.

Analysis

While training the neural network if data reached to the point of validation check, it will break down the network and stops it. According to hypothesis if decrement occurs in the value of validation check than there will be absolute decrease in the accuracy. According to experiments performed by changing its values it did not hit back the accuracy significantly. In some cases when there was a big change occurred in the values than accuracy was little bit moved according to hypothesis.

References

Mehdy, M., Ng, P., Shair, E., Saleh, N. and Gomes, C. (n.p.). Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer.

Christina Chun, M. (n.p.). Breast cancer: Symptoms, risk factors, and treatment. [online] Medical News Today. Available at: <https://www.medicalnewstoday.com/articles/37136.php> [Accessed 3 Dec. 2017].

Saritas, I. (2012). Prediction of Breast Cancer Using Artificial Neural Networks.