2017

Breast Cancer Classification Using Neural Network Approach

EHTISHAM UL HASSAN MALIK
14031186

03-12-2017 NEURAL NETWORKS AND FUZZY SYSTEMS

ABSTRACT

Pollution is increasing day by day globally and by that the incidence and severity of diseases has also been increased as the climate change models predicted. Cancer is one of the deadliest disease and thousands of its types have been discovered. People are even referring twentieth century as Cancer century. Breast cancer is one type of cancer which is found in the women. Thousands of women have died in recent years suffering from this disease. Scientists are working very hard to find the perfect cure of this disease. Among these scientists are computer scientist who are working on neural networks to make sure that patient is diagnosed in early stages of breast cancer so it can be easily cured.

This paper is written for a test of neural network on a data set which contains the information about the breast cancer patients. A lot of work has already been done on these kind of experiments. This paper can be helpful in improving the previous work.

CONTENTS

INTRODUCTION

A tumor or cancer is by definition a neoplastic process where abnormal cellular growth takes place and in which all the newly formed tumor cells are descendants of a single cell that gained the ability to replicate continuously and autonomously (Alison pt. 3; ch. 39).

Breast cancer is the most frequent cancer among women, impacting over 1.5 million women each year, and also causes the greatest number of cancer-related deaths among women. In 2015, 570,000 women died from breast cancer – that is approximately 15% of all cancer deaths among women. While breast cancer rates are higher among women in more developed regions, rates are increasing in nearly every region globally. In order to improve breast cancer outcomes and survival, early detection is critical. (World Health Organization, 2017)

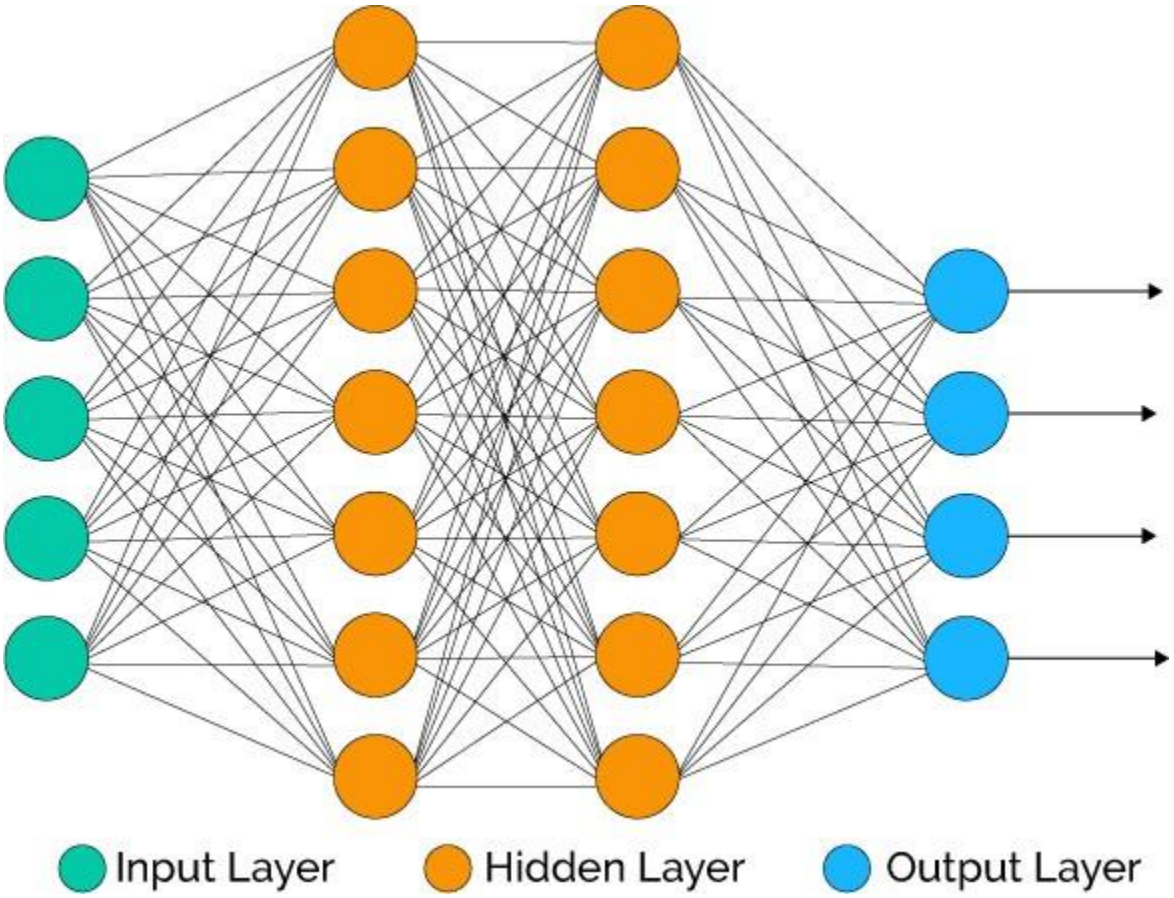
1.67 million cases of breast cancer were discovered in 2012 all over the world. Although this diseases is only found in women but it is still 5th largest cause of deaths when it comes to diseases. (Globocan.iarc.fr, n.d.)

To cure breast cancer it is necessary that it is diagnosed in early stages. Therefore, researchers nowadays are using neural networks to help them diagnose the patient. I have designed a neural network which is trained on the set of actual data not on some rules. The data set contain the data of almost 600 patients and I am using half of the data to train the data and half of it to test the neural network.

BACKGROUND AND RELATED WORK

Neural Networks is very vast and interesting topic in computer sciences. This is because it resembles with the human brain as it also comprises with trillions of neurons. The accuracy of the neural networks can be as good as human brain in most of the cases. Neural Network is the organized structure of neurons which are interconnected with each other forming a complex structure. These complex structures are responsible for receiving data and sending signal into the system. These signals are controlled by the activation function

which decides which direction these signals should go. Not only direction but the magnitude can also be defined. Neural network are composed of three layers: input, output and hidden layer. Data is received by first layer i.e the input layer while the last layer, as can be predicted, sends the data out so it is called output layer. Between these two layers is the hidden layer which does all the work and its structure is very complex.



Sciences and Communication in 2012 demonstrated truly great outcomes. The analysts utilized "Wisconsin informational collection" having 699 examples with 16 having missing esteems. The information is isolated in two primary parts, that is, preparing and testing information. Among this data, 80% of the information comprises of training and 20% is testing data. Different neural systems have been attempted keeping in mind the end goal to contrast the outcomes and each other and to search for the best one regarding comes about. While utilizing Radial Base Function (BRF) procedure which has three layers and non-straight change from contribution to direct and direct change from covered up to yield. The precision gained with the assistance of this method was 96.18%. When they utilized Probabilistic Neural Networks (PNN) with Bayes' hypothesis alongside Parzen Windows's strategy the precision went up to 97%. The utilization of General Regression

Neural Networks (RNN) expanded the exactness to 98.18%. The most astounding precision was accomplished with the assistance of a strategy named as Back Propagation Neural Network (BPNN) and the exactness accomplished was 99.28%. (Swathi et al. 2012)

Another technique Feed Forward Neural Networks (FFNN) along with back propagation algorithms was used to diagnose breast cancer. Pawar and Patil used confusion matric for the system evaluation. They also used the same data set. Wisconsin data set was chosen to be used in this system for training and testing purposes. This data set contained total number of 699 records 19 of which were not right and were later excluded from the dataset. The remaining data contain 458 records of benign and 241 records of malignant. After training their system they got 99% accuracy using nine hidden layers, 1000 epochs. The network used in this system was Radial Base Function (RBF).

METHODOLOGY

Data Preprocessing

The "Wisconsin data set" contains 11 columns and 699 rows. The first column of the data set is the id of the patient. The columns from 2 to 10 contains the data about the patient while the last column contains the diagnoses result of the patient. The result in the data set is of two types: benign and malignant. So in order to train the neural network more efficiently I sorted the given data as one malignant and one benign i.e if first row is benign than the second will be malignant and the third will be benign and so on. After that the data is further separated into two parts: the input data which contains the columns from 2 to 10 and the output which contains the last column of Wisconsin data set. Some data from the input was missing as there was '?' in place of data so I removed that data and replaced it with the mean of 6^{th} column.

Hypothesis1

My hypothesis was that if we use more data for training and less data for testing the results will be more accurate. Therefore, I took 80% of the data for training and 20% of the data for the testing. As for learning rate and no of neurons my idea was that if we keep increasing them with each test the performance of the system will get better.

Results

In the first experiment the results were pretty good but not as good as expected. The arrangements of the dataset used and the other factor resulted in achieving the 78% accuracy of the system.

Training Data	**Testing Data**	**Learning Rate**	**Neurons**	**Accuracy**
80%	20%	0.01	10	78.4170%

Hypothesis2

After the network was trained on the 80% of the data and tested for 20% of the data. I thought it might be a good idea to take less data for training and more for testing. So I decided to take 60% of the data from dataset for training and 40% for the training purposes.

Results

In the first experiment the results were pretty good but not as good as expected. But the result of this experiment were amazing. With the learning rate 0.05 and 15 no of neurons used the system gave the accuracy of 87%

Training Data	**Testing Data**	**Learning Rate**	**Neurons**	**Accuracy**
60%	40%	0.05	15	87.102%

Hypothesis3

The results of the previous experiment were good. So I decided to make it little bit more challenging for the system. I divided the data into 50-50. Half of the dataset for training and half for testing.

Results

This experiment also gave promising results. The arranged data set with the learning rate of 0.08 and 20 number of neurons gave the accuracy of 86%

Training Data	**Testing Data**	**Learning Rate**	**Neurons**	**Accuracy**
50%	50%	0.08	20	86.5766%

Hypothesis4

The results of the previous experiment were good. So I decided to make it little bit more challenging for the system. I now take 40 % of the data as testing data and 60% for the testing.

Results

This experiment also gave promising results. The arranged data set with the learning rate of 0.08 and 20 number of neurons gave the accuracy of 87%

Training Data	Testing Data	Learning Rate	Neurons	Accuracy
50%	50%	0.1	25	87.3258%

Hypothesis5

The results of the previous experiment were good. This is the last experiment and should be difficult of all. Therefore the 20% of the dataset will be used for training and remaining will be used for testing.

Results

The results of this experiment was as good as expected. This can be the result of number of neurons used as they are making system more complex or maybe the system was not trained enough to test such huge amount of data.

Training Data	Testing Data	Learning Rate	Neurons	Accuracy
20%	80%	0.5	30	78.8576%

CONCLUSION

Total number of five experiments were done with the different amount of the data with slightly changed structure of the system itself. It was observed that all the experiments gave promising results except the first one. But the accuracy of all the experiments was almost 80% which shows that the system is working efficiently. Although the 100% of the accuracy was not achieved during this process which can be justified because there is always room for error even if we talk about human beings. Or maybe the system was not complex enough but as we have seen that if we increase the number of neuron above certain level the system become more complex and get confused and the accuracy of the system decreases. From these experiments it is observed if we design the system more precise and use it more properly good results can be obtained and not only breast cancer but lot of other diseases can also be diagnosed using neural networks.

REFERENCES

Alison, Malcolm. The Cancer Handbook. 2nd ed. Chichester, West Sussex, England: John Wiley & Sons, 2007. Print.

Pawar and D. R. Patil. Breast Cancer Detection Using Neural Network Models. pages 568- 572, 2013.
World Health Organization. (2017). *Breast cancer.* [online] Available at: http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.[Accessed 1 Dec. 2017].

Globocan.iarc.fr.(n.d.).Fact sheet by cancer. Available at: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx [Accessed 1 Dec. 2017].

Swathi, S., Rizwana, S., Babu, A., Kumar, P. and Sarma, P. (2012). *CLASSIFICATION OF NEURAL NETWORK STRUCTURES FOR BREAST CANCER DIAGNOSIS*. Available at: http://www.csjournals.com/IJCSC/PDF3-1/Article_48.pdf [Accessed 1 Dec. 2017].