BREAST CANCER RISK AND DIAGNOSTICS USING ARTIFICIAL NEURAL NETWORK(ANN)

Vraj Shah[1]

[1]B.Tech, Computer Science and Engineering, MIT School of Engineering, Pune, Maharashtra, India

---

Abstract – Breast Cancer is one of the serious disease that causes high number of deaths every year.It is one of the common cancers and prominent in death of women worldwide.Using Artificial Neural Network we will try to classify if the cancer is Benign or Malignant.Various performance parametres are used in the model on the basis of which the model is evaluated . The main objective of the model is to generate accurate results on the basis of input .The Artificial Neural Network is implemented on the Wiscoin Breast Cancer dataset from the UCI Machine Learning Repository.

Key Words – ANN, RMSE, MEA, LOSS, ACCURACY, KAPPA STATISTICS, MACINE LEARNING

1. Intoduction – Breast cancer is cancer that forms in the cells of the breasts. Breast cancer can occur in both men and women, but it's more common in women. Breast Cancer is the major cause of death of women after Lung Cancer.

Substantial support for breast cancer awareness and research has helped created advances in the diagnosis and treatment of breast cancer. Breast cancer rate of have increased, and the number of deaths due to this disease is steadily declining, largely due to factors such as earlier detection, a new personalized approach to treatment and a better understanding of the disease.

Using these models we can do our data analysis. This can help to predict the results and help to reduce the cost of medicines make lives of people better and make real time decisions. There are many algorithms for prediction of breast cancer outcomes. This paper gives the prediction based on the artificial neural network developed which is an important part of machine learning. This model is evaluated on 683 instances where 0 is benign and 1 is malignant.

This model is create using 2 layered neural network. After execution it is evaluated on the basis of root mean square method (RMSE),mean absolute method(MEA),kappa statistics and it produces accuracy of 97%.The main aim is to evaluate efficiency and effectiveness of algorithm in terms of accuracy and precision

2. Related Work - Classification is one of the most important and essential task in data mining and machine learning. A lot of research has been done to apply data mining and machine learning on different medical datasets to classify Breast Cancer. Many of them show good classification accuracy.

Vikas Chaurasia and Saurabh Pal[11] compare the performance criterion of supervised learning classifiers; such as Naïve Bayes, SVM-RBF kernel, RBF neural networks, Decision trees (J48) and simple CART; to find the best classifier in breast cancer datasets. The experimental result shows that SVM-RBF kernel is more accurate than other classifiers; it scores accuracy of 96.84% in Wisconsin Breast Cancer (original) datasets.

V Nandakishore[13], compare the performance of C4.5, Naïve Bayes, Support Vector Machine (SVM) and K-Nearest

Neighbor (K-NN) to find the best classifier in WBC. SVM proves to be the most accurate classifier with accuracy of

96.99%. Angeline Christobel. Y and Dr. Sivaprakasam[14], achieve accuracy of 69.23% using decision tree classifier (CART) in breast cancer datasets.

3. EXPERIMENT – The experiment is conducted using python language and spyder ide. Machine Learning libraries like sci-kit learn, keras etc are used for executing this model

3.1 Artificial Neural Network-

Artificial neural network is a computational model that works like a human brain.It consists of

Neurons which are connected to each other and trained on the given set of data to produce output.

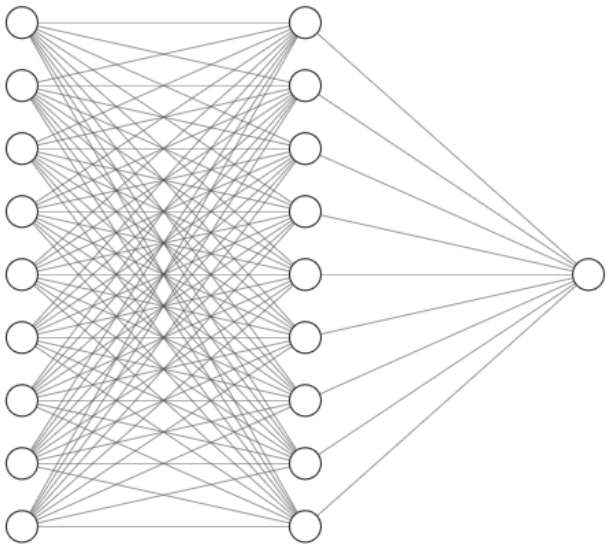

Fig 1.Neural Network

The model uses the equation given below-

$$y = \sum_{i=1}^{n} (ai * wi) + b$$

Where, y=output

a=input to neuron

w=weight

b=bias

3.2 Effectiveness – In this section we will evaluate the effectiveness of the model by classifying correct predictions, incorrect predictions and accuracy.

Table 1.Train Test Data

	Train	Test
Correct Prediction	534	134
Incorrect Prediction	12	3
Accuracy	97.8%	97.5%

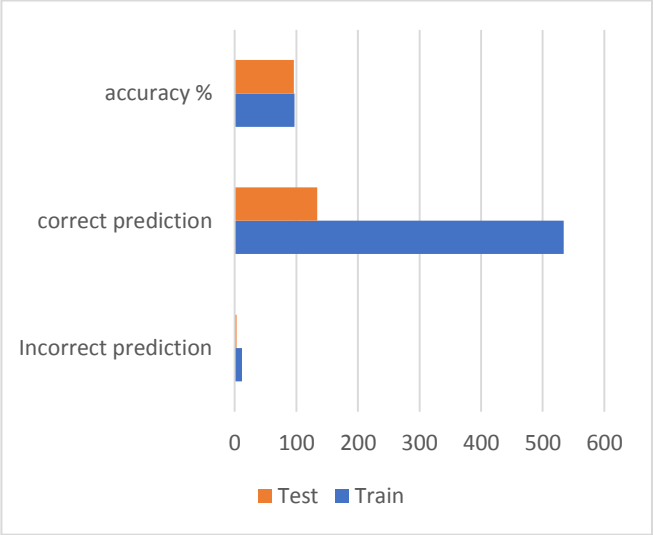

Fig 2.Accuracy Score

The accuracy score is evaluated from the confusion matrix which was implemented on the test and train data on the total instances of 683.

To find efficiency of this model is evaluated by taking into consideration various parameters such as mea,rmse,kp.The purpose of error measurement is to find more robust summary of the distribution.The common practice of calculating error is by first calculating the loss function and then computing the average.

[1] MEA- Mean absolute error in statistics is a measure of errors between paired observations expressing the same phenomenon.

$$MEA = \frac{\sum_{i=1}^{n} |yi - xi|}{n}$$

[2] RMSE – Root mean square error is used

frequently to measure the differences between values (sample or population values) predicted by a model or an estimator and the values observed.

$$RMSE = \sqrt{MEA}$$

[3] Kappa Statistics – Kappa statistics is a chance corrected measure of aggrement between the classifications and the true classes.

The fig below gives a graphical view of the comparison of MEA,RMSE and KS for the train and test data.

Table 2. Error Measurement

	Train	Test
MEA	0.0219	0.0218
RMSE	0.1482	0.1479
KS	0.9516	0.9533

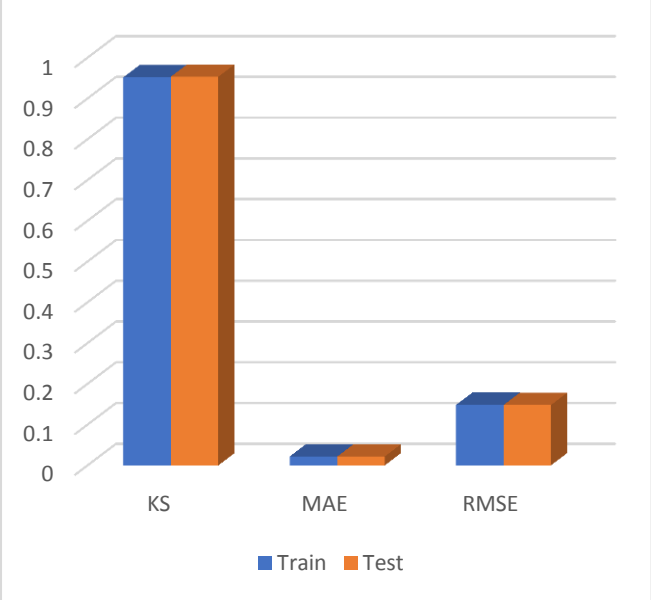

Fig 3. Comparison between evaluation parameters

3.3 Efficiency- We will use model loss and model accuracy to evaluate how efficient the model is implemented to find out whether the modl is overfitted or not.A machine learning algorithm can be evaluated using a loss function . The loss function is calculated on training and validation and its interpretation tells us how good the model is. It is the sum of errors for example in training or validation sets.Loss value tells us how poorly or well a model behaves after each iteration of optimization.

An accuracy metric is used to measure the performanceof the algorithm. The accuracy of a model is determined usually after the model parameters and is calculated in the form of a percentage.It basically tells us how accurate our model's prediction is as compared to the true data.

The stochasticstic gradient decent method is used to calculate the weights that will give us the best results and minimum error.

	Train	Test
Benign	357	87
Malignant	189	50

Table 3.Total Benign and Malignant

The figs below below show the accuracy and the loss for the data at different epochs.

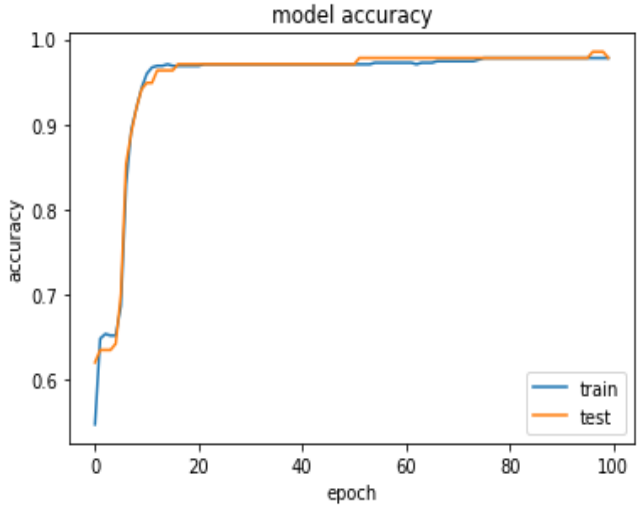

Fig 4. Accuracy

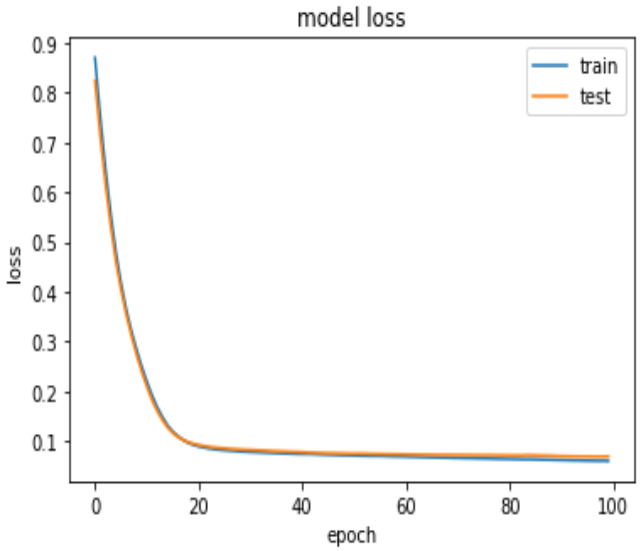

Fig 5.Loss

From the above fig we can see that model seems to be a good fit as the difference between the accuracies between the training and test set at different epochs is not much.The no.of wrong predictions in the train set for benign are 8 and for malignant it is 4. Whereas in the test set the wrong predictions for the benign are 3 and for malignant it is 0.

4. Conclusion – To do data analysis various mining and machine algorithms are available. But the challenge is to build a accurate models which can give accurate results.The ANN model build gives an accuracy of 97.8% which outperforms the models dicussed in the above section.To conclude ANN model has proven its efficiency in Breast Cancer prediction and Diagnosis and also achieved best performance in terms of precision and low error rate.

5. References –

1. "UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29. [Accessed: 29-Dec-2015].
2. S. Aruna and L. V Nandakishore, "KNOWLEDGE B ASED A NALYSIS OF V ARIOUS S TATISTICAL T OOLS IN D ETECTING B REAST," pp. 37–45, 2011.
3. A. Pradesh, "Analysis of Feature Selection with Classification : Breast Cancer Datasets," Indian J. Comput. Sci. Eng., vol. 2, no. 5, pp. 756–763, 2011.
4. Djebbari, A., Liu, Z., Phan, S., AND Famili, F. International journal of computational biology and drug design (ijcbdd). 21st
5. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2012.