Deep Neural Network Classifier in Breast Cancer Prediction

S. Poorani, P. Balasubramanie

Abstract*: The Breast cancer is the most life menacing disease among women. Early prophecy assurances the endurance of patients. In this work, first Deep neural network classifiers with different hidden layers with different nodes are used to explore the anthropometric information and blood investigation strictures and to predict the disease. Then machine learning algorithms such as SVM and Decision tree are also trained with the same data. Finally the performance of each classifier was deliberated. The pre-processed data of admitted patients with the breast cancer perception are used to train and test the classifiers. This article shack glow on the concert estimation based on right and erroneous data classification

Keywords* : *Breast Cancer, Classification, Deep Learning, Neural Network .

I. INTRODUCTION

Deep neural network is the evolving technique that can be applied in a variety of applications like pattern recognition, natural language processing, medical image, bio informatics, audio recognition and so on. It is also successful in clinical data classification. It is suitable for non-linear dataset with more number of inputs. In classifying the cancer dataset Neural network was the popular tool because, it has the ability to characterize the complex and multidimensional linear functions as well as non-linear functions[3]. The key advantage of neural network is that the models are trained with examples not the rules. Multilayer Perceptron is the suitable method to envisage the survival of breast-cancer syndrome to acquire improved outcome [6]. To improve accuracy in breast-cancer prediction several supervised algorithms like SVM, Decision tree, ANN and K-nearest neighbor can be interpreted. Unsupervised algorithm like clustering and semi-supervised methods can also be infered[7].

The technologies make the researchers to avail the datasets easily from various domains from the centralized and open-accessed repositories. The UCI-Repository is such a repository where number of datasets are available including Breast-cancer data. The most commonly used datasets from this library are BCWD, WDBC, WPBC and Coimbra [3]-[10]. Generally accuracy is used as main performance measure, so that we find and compare the accuracy of three classifiers(DNNs, SVM & Decision tree) with the Coimbra dataset.

.

Revised Manuscript Received on October 15, 2019.

S. Poorani, Assistant Professor, Kongu Engineering College, Perundurai, Erode, Tamilnadu, India. Email: vspoorani@gmail.com

P. Balasubramanie, Professor, Kongu Engineering College, Perundurai, Erode, Tamilnadu, India. Email: pbalu_20032001@yahoo.co.in

II. RELATED WORKS

Harikrishnan et al[2] applied the neural network model to analyze the classification of satellite images. They used back propagation algorithm with ensemble methods and radial-basis function. In which the model with ensemble methods provide better accuracy. Htet Thazin Tike Thein and Khin Mo Mo Tun [3] used the WBCD database to test the concert of neural network. It identifies the type of breast cancer as malignant or benign. The dataset contains nine predictors and one classifier. They used the island propagation algorithm with different topologies. Muchid Mustafa Saritas, Ali Yasar[4] applied the ANN method and Naivebayes method to classify the breast-cancer(Coimbra dataset), in which ANN provides higher accuracy than Naivebayes. Chandra Prasetyo Utomo et al[5] implemented the extreme-learning method in ANN to diagnose the dataset named Breast-Cancer Wisconsin (BCWD). It provides healthier concert than Back-propagation method. A. Kathija et al[6] used three datasets from UCI library named WBCD, WDBC and WPBC to evaluate the performance of MLP model. S. Roobini and J. Fenila Naomi aicenisha[7] clustered the WBCD with fuzzy c-means method and also investigated ANN, SVM and Naïve bayes methods with the clustered data. Subrata Kumar Mandal[8] measured the time complexity of logistic regression, naïve bayes and decision tree methods with WDBC, in which Logistic regression shows better performance. Hiba Asria et al[9] employed the KNN,SVM,C4.5 and NB with WBCD in which SVM shows better performance. R.R.Janghel et al[10] implemented four nn models with WBCD. They are competitive learning method, back propagation, radial-basis function and Linear vector-quantization methods. The LVQ shows high accuracy. In the literature the authors analyzed the recitation of nn classifier with data from various domains like satellite images to breast cancer dataset. Four different types of breast cancer datasets were analyzed in the previous works. They are WBCD, WPBC, BCWD and Breast-cancer Coimbra dataset from UCI library[3], [4], [5], [6], [7], [8], [9], [10]. Even-though the breast-cancer dataset was analyzed with nn method, no more than one hidden layer was employed. In our work we exploit the Coimbra breast-cancer dataset from UCI library[1] to train the Deep Neural-network classifier.

III. METHODOLOGY

A. Dataset

The coimbra-dataset[2] used in this experimental-setup was taken from the UCI library. It is a multivariate dataset with 116 observations and 10 features.

Among 116 instances 64 instances are about breast cancer patients and remaining 52 are about healthy controls. Among ten attributes nine are independent variables(age, BMI, Glucose, Insulin, HOMA, Leptin, Adeponectin, Resistin, MCP-1) and the 10th variable is the dependent variable which has the label 2 to represent presence of the disease and 1 to represent the absence of the disease.

B. Classification

Deep neural network is a neural network which can have two or more hidden layers. In this work we tried five DNN classifiers with tanh activation function and back propagation algorithm. The same dataset[2] is used in all the five models. 60% of the data were used to train the model and the remaining 40% were used to evaluate the classifier performance. The thresold value is 0.01 and the stepmax value is 1e+05. The networks with different hidden layers and different nodes are shown in the Fig1,Fig 2 Fig 3, Fig 4 & Fig 5. The same train-data and test-data is used with machine learning models such as SVM and Decision tree methods.

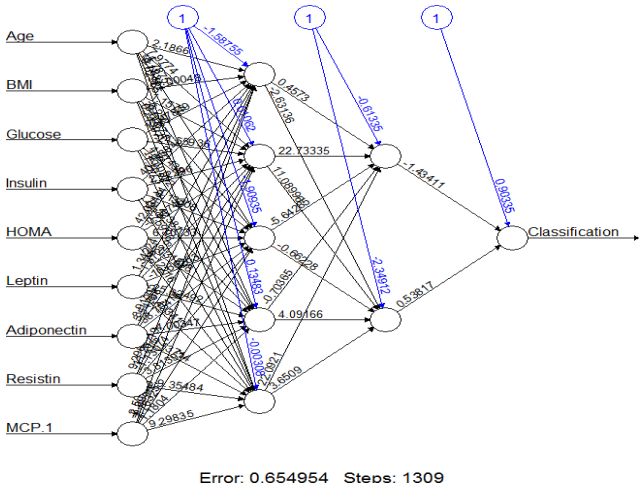

Fig. 1 Deep Neural Network with Two hidden layers with 5 and 2 nodes.

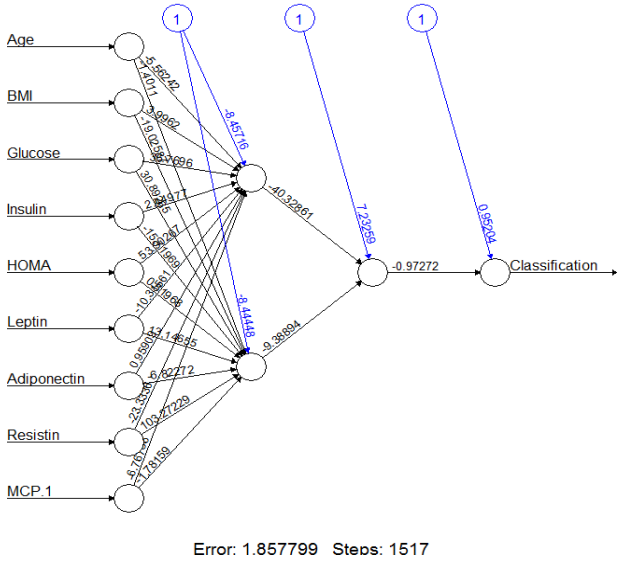

Fig. 2 Deep Neural Network with Two hidden layers with 2 and 1 node.

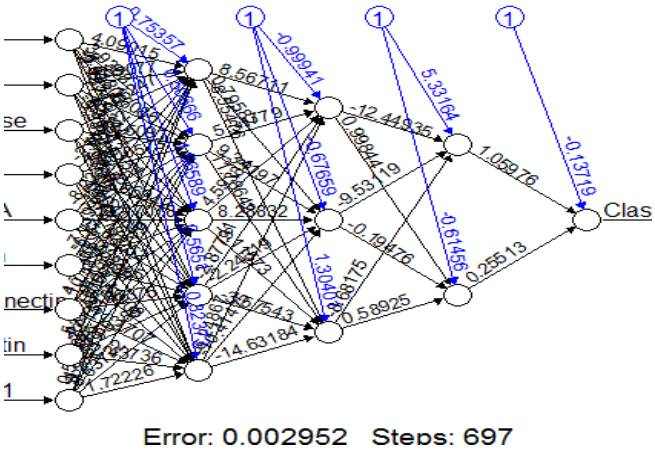

Fig. 3 Deep Neural Network with Three hidden layers with 5, 3 and 2 nodes.

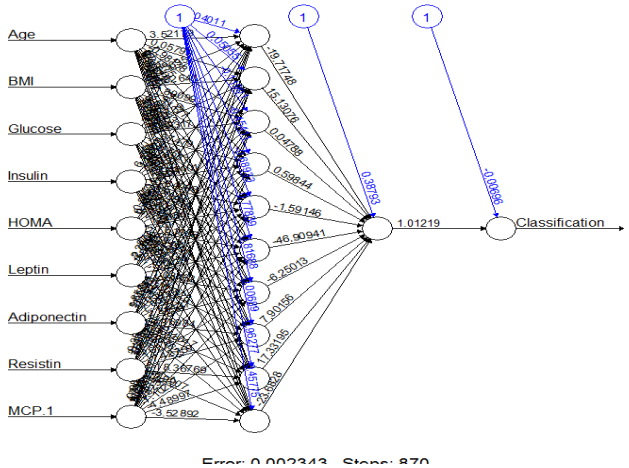

Fig. 3 Deep Neural Network with Two hidden layers with 10 and 1 node.

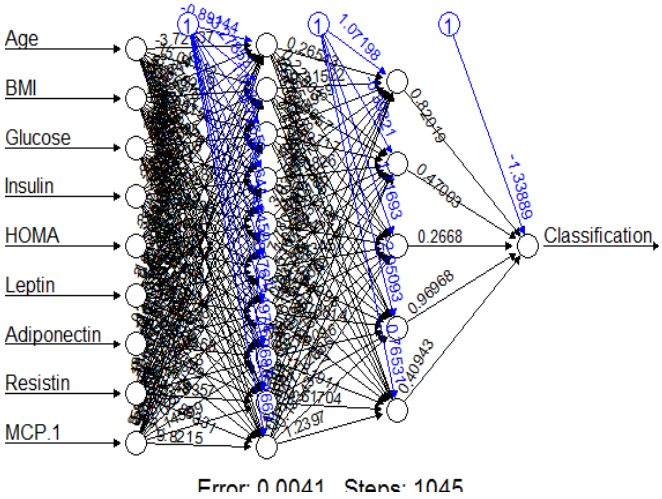

Fig. 5 Deep Neural Network with two hidden layers with 10 and 5 node.

IV. RESULTS AND DISCUSSION

The confusion matrix and classification performance of each DNN model is given in table I and table II respectively. From table II the result shows that network which has 2 hidden layers with 5 nodes and 2 nodes provides the higher accuracy.

The network with three hidden layers provides lower accuracy than the classifiers with two hidden layers. The accuracy value of other machine learning methods are given in table III. While comparing to SVM and Decision tree DNN provides better accuracy.

V. CONCLUSION

The DNN classifiers provide good accuracy for small dataset also. They will be more efficient in large datasets with more number of attributes and may provide better accuracy. Thus we can use these types of classifiers for other clinical datasets with large datasets or big data.

Table III: Classification Performance Of Other Classifiers

Classifier	Accuracy(%)
SVM	65.96
DECISION TREE	40.83

Table- I: Confusion Matrix of Neural Network Classifier

No. of hidden layers	No. of nodes in each layer	Actual	Prediction	
			0	1
2	Layer-1: 5 nodes Layer-2: 2 node	0 1	12 3	9 23
2	Layer-1: 2 nodes Layer-2: 1 node	0 1	11 5	10 21
3	Layer-1: 5 nodes Layer-2: 3 node Layer-3: 1 node	0 1	8 5	13 21
2	Layer-1: 10 nodes Layer-2: 1 node	0 1	12 8	9 18
2	Layer-1: 10 nodes Layer-2: 5 node	0 1	12 6	9 20

Table- II: Classification Performance Of DNN

Classifier	Accuracy(%)
DNN with 2 hidden layers with 5 nodes and 2 nodes respectively	75.94
DNN with 2 hidden layers with 2 nodes and 1 node respectively	68.25
DNN with 3 hidden layers with 5nodes,3nodes and one node respectively	61.65
DNN with 2 hidden layers with 10 nodes and 1 node respectively	63.33
DNN with 2 hidden layers with 10 nodes and 5 nodes respectively	67.82

REFERENCES

1. [Patricio, 2018] Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seiça, R., & Caramelo, F. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer, 18(1), https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra.
2. Harikrishnan R, Shivali Amit Wagle, "Analysis of Classified Satellite Images Using Different Neural Networks", International Journal of Engineering and Advanced Technology. Vol-8, Issue-5, June 2019, 379-382.
3. Htet Thazin Tike Thein, Khin Mo Mo Tun, "An Approach for Breast Cancer Diagnosis Classification Using Neural Network",. Advanced Computing: An International Journal, Vol-6, Issue-1, Jan 201,pp. 1-11.
4. Muchid Mustafa Saritas, Ali Yasar, "Performance Analysis and Naive Bayes Classification Algorithm for Data Classification", International Journal of Intelligent Systems, Vol-7 No-2, 2019, pp. 88-91.
5. Chandra Prasetyo Utomo, Aan Kardiana, Rika Yuliwulandari, "Breast Cancer Diagnosis using Artificial Neural networks with Extreme Learning Techniques", International Journal of Advanced Research in Artificial Intelligence, Vol-3 No-7, 2014,pp. 10-14.
6. A. Kathija, S. Shajun Nisha and Dr .M .Mohamed Sathik, "Breast Cancer Data Classification Using Neural Network Approach of MLP Algorithm", International Journal of Trend in Research and Development, Vo-4 No-3, May-June 2017,pp. 275-279.
7. S. Roobini and J. Fenila Naomi aicenisha, "Performance Analysis of Different Classifiers in Prediction of Breast Cancer", Indian Journal of Science and Technology, Vol-12 No-8, February 2019, pp.1-6.
8. Subrata Kumar Mandal, "Predict Performance Analysis Of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree", International Journal Engineering and Computer Science, Vol-6 No-2, February 2019, pp. 20388 – 20391.
9. Hiba Asria, Hajar Mousannifb, Hassan Al Moatassimec, Thomas Noeld. "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis", The 6th International Symposium on Frontiers in Ambient and Mobile Systems. Procedia Computer Science 83, 2016, pp. 1064-1069.
10. 10. R.R.Janghel, Anupam Shukla , Ritu Tiwari , Rahul Kala, "Breast Cancer Diagnosis using Artificial NeuralNetworkModels" ,https://www.researchgate.net/publication/224163164_Breast_Cancer_diagnosis_using_Artificial_Neural_Network_models, March 2019, pp. 89-94.

AUTHORS PROFILE

Ms. S. Poorani, received M.Sc degree at Sri Vasavi College, Bharathiar University from the Department of Computer Science, India, in 2004. She has 13 years of teaching experience. She published 8 articles in International journals. She has also presented papers in National Conferences and International Conferences. Her area of interest includes data mining and big data. She is currently pursuing the Ph.D. degree working with Dr. P. Balasubramanie. Simultaneously she is also currently working as an Assistant Professor in the department of Computer Technology, Kongu Engineering College, affiliated to Anna University, Tamilnadu, India.

Dr.P.Balasubramanie, is currently working as a Professor in the Department of Computer Science and Engineering, Kongu Engineering College, Tamilnadu, India. He is one of the approved Supervisor of Anna University, Chennai and guided 26 Ph.D Scholars. Currently he is Guiding 11 scholars. He has published 218 articles in International/National journals. He has authored/co-authored 11 books with the reputed publishers. Three of the books published are used as text/reference books by many of the leading Universities in India. He has completed one AICTE RPS as a Principal investigator and currently he is working as a principal investigator in a MRP of UGC. He has received Rs. 13 Lakhs of grant from various funding Agencies like AICTE, CSIR, NBHM, DRDO, INSA and so on and organized 21 STTP/SDP/Seminar/Workshops for the benefit of Faculty members and Research scholars. He has received 17 awards so far from various agencies. His area of interest includes Data mining, Networking, cloud computing and Optimization algorithms.