ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 9, Issue, 4(D), pp. 25788-25792, April, 2018

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

BREAST CANCER DATA CLASSIFICATION USING NEURAL NETWORK AND DEEP NEURAL NETWORK TECHNIQUES

Arpita Joshi[1] and Ashish Mehta[2]

[1]Department of Computer Science, Kumaun University, S.S.J. Campus, Almora
[2]Department of Computer Science, Kumaun University, D.S.B. Campus, Nainital

DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.1931

ARTICLE INFO

Article History:

Received 8th January, 2018
Received in revised form 21st February, 2018
Accepted 05th March, 2018
Published online 28th April, 2018

Key Words:

Machine learning techniques, Breast Cancer dataset, R Programming Language.

ABSTRACT

Breast cancer is one of the most common reasons for cancer death. Early detection and diagnosis are very important for the survival of the patient. There are various Machine Learning techniques available for the purpose of diagnosis of Breast Cancer. In this study, we have compared the classification results of Neural Network and Deep Neural Network Technique for classifying Breast Cancer dataset into benign and malignant. In the present study we analyzed the classification results of Neural Network and Deep Neural Network techniques with and without dimensionality reduction techniques (Principal Component Analysis and Linear Discriminant Analysis).
We took Wisconsin Diagnostic Breast Cancer dataset available from UCI repository containing 32 attributes. The performance of each technique is evaluated using various performance measures. The classification result shows that Neural Network technique and Deep Learning technique both gives better result in terms of different parameters.

INTRODUCTION

In the present era, cancer is one of the most common diseases responsible for multitude human deaths. The only chance to fight cancer is to detect it in its early stage. There are different types of cancers occur in the human body. Breast cancer is the second leading cause of death in women [7]. Breast cancer occurs in male and female both as well as in animals [16]. Breast cancer diagnosis and prognosis depends on various factors including the type of cancer, stages, treatment etc [12]. According to the report of National Institute of Environmental Health Sciences, Approximately 246,000 women are diagnosed with breast cancer in the United States [11]. According to the report available at www.cancer.org, there are various approaches to reduce breast cancer risks including hormone therapy drugs, dietary supplements. Breast cancer risks will increase with age. In the United States, the death rates from breast cancer mortality have declined 39% between 1989 and 2015[9].Breast cancer starts when cells in breast begin to grow out of control and form a tumor. Breast cancer spreads when cancer cells get into blood or lymph and carried out to different parts of the body. Most of the breast lumps are not cancerous (benign), they are abnormal growths but they do not spread outside of the breast and consequently are not life threatening. If the cell grows into invading tissues, it becomes a malignant tumor. In other words, we can say that Breast cancer can be detected with tumor classification Breast cancer starts from different parts of the breast. There are two types of breast cancer: ductal cancer and lobular cancer. According to the American Cancer Society and the National Cancer Institute, in the US Every year approximately 250,000 new cases will be diagnosed in female and over 2,400 in men [9, 10, 12, 18].

Early Detection of breast cancer is very important for the correct diagnosis. There are various Machine Learning techniques available for the purpose of diagnosis of breast cancer. In the present study, we have employed Neural Network and Deep Neural Network techniques from the domain of Machine Learning tools for classification of breast cancer into benign and malignant. Section 2 gives a brief literature survey. The proposed system model is being presented in section 3. Section 4 showcases experimental results. Finally, concluding remarks are drawn in section 4.

A brief literature review

Karabatak and Ince (2009) experimented on Wisconsin breast cancer dataset (699 records and 9 attributes) and applied 3-Fold Cross Validation method in the test stage. They used two Association Rules (AR1 and AR2) and concluded that the combination of Association Rule 1 and Neural Network gives

Corresponding author:* **Arpita Joshi
Department of Computer Science, Kumaun University, S.S.J. Campus, Almora

the best classification performance with the classification rate of 97.4% [15].Mirza and Jain reached 96.44% testing accuracy and 91% training accuracy using Back Propagation Neural Network [4] .Baoyu Zheng Wei Qian, and Laurence P. Clarke (1996) employed Back Propagation method with Kalman Filtering. They used mixed Feature Neural Network technique and calculated the True Positive (TP) and False Positive (FP) values for both raw (TP-.814, FP-.59) and enhanced images (TP-.901, FP-.71)[3]. Chen Ren Dar, Chang Feng Ruey and Huang Len Yu (2000) reached 85.6% classification accuracy using 10 fold cross validation with Self Organizing Map method of Neural Network[6]. Jihong Liu and Weina Ma (2007) resulted that Support Vector Machine (96.12%) gives better result in comparison to Back Propagation Neural Network (93.33%) [13]. Endo A., Shibata T. and Tanaka H. (2008) analyzed various machine learning techniques to predict survival rate for breast cancer patients. They used data from the SEER program and found that Artificial Neural Network had the highest specificity [8]. Yan Xu,Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai and Eric I-Chao Chang. (2014) investigated a Deep Learning Technique through which they obtained automatic extraction of feature representation. The performance of unsupervised feature learning framework was 93.56%, fully supervised feature learning framework reached 94.52%, and the last framework supervised deep learning feature with multiple instance learning obtained 96.30 % [20].Ahmed M. Abdel Zaher and Ayman M. Eldeib (2015) used Deep Belief Network with Back Propagation Neural Network to classify breast cancer into benign and malignant. The accuracy obtained by them was found to be 99.68% [1].

Proposed System Model

There are various Machine Learning techniques available for the purpose of diagnosis of breast cancer into benign and malignant. In the present study, we use Neural Network and Deep Neural Network Techniques for the detection of benign and malignant. We employed *R programming language* for this purpose.

Proposed System Model

The data set available in the UCI Machine Learning Repository will be used for this study. The data will be preprocessed first and then different Machine Learning techniques such as Neural Network and Deep Neural Network will be applied on that data which consequently generates different features that can be associated with the diagnosis of breast cancer.In order to evaluate the performance of Machine Learning techniques, various performance measures like accuracy, precision, sensitivity, and specificity etc. are used**[5,14].** An appropriate model for the classification will be proposed and validated. There are various steps involved in the proposed system model.

Data Collection

The data will be collected from UCI Machine Learning Repository to differentiate between malignant and benign **[5].** A brief description of the dataset is presented in table 3.1.1 below:

Table 1 Breast Cancer Data Set

Dataset	Attributes	Instances	Classes
Breast Cancer	32	569	2

Data Preprocessing

Data preprocessing is the first and thus very important step. The quality of the data and the amount of useful information that it consists are key factors which determine how adequately a machine learning technique can learn. Most computational tools are unable to deal with missing values, or they produce unpredictable outcomes if we simply ignore them. One of the easiest ways to deal with missing data is to remove the corresponding columns (features) or rows (samples) from the dataset entirely. The removal of rows or eliminating entire columns is not feasible as this may lead to loss of most valuable data. For this reason, we use one of the interpolation technique *mean imputation* to estimate the missing values from the other training samples [17].

Feature Scaling

Feature Scaling is one of the most crucial steps in preprocessing. There are two important techniques to bring different features on to the same scale [17].

1. Normalization: Rescaling of the capabilities to a range of [0, 1].
2. Standardization: Using this technique, we are able to center the feature columns at mean zero with standard deviation 1.

Dimensionality Reduction

There are different methods for reducing the dimensionality of a dataset using different Feature Selection techniques. One of the most common techniques for Feature Selection for Dimensionality Reduction is Feature Extraction. Feature Extraction is a method that transforms the data into a new feature space [17]. There are mainly two types of Dimensionality Reduction techniques:

Principal Component Analysis (PCA): Principal Component Analysis is an unsupervised linear transformation technique. PCA is used for feature Extraction and Dimensionality reduction.PCA facilitates to discover patterns in data based on the correlation between features. If we use PCA for Dimensionality Reduction, construct a d*k dimensional transformation matrix W that allows mapping a pattern vector x on to a new k-dimensional feature subspace [17].

$X=[x_1, x2, x3, xd_]$, x belongs to R^d
XW, W belongs to R^{d*k}

Linear Discriminant Analysis (LDA): Linear Discriminant Analysis is supervised data compression technique. Linear Discriminant Analysis is an important method for Feature Extraction to increase the computational efficiency and reduce the degree of Overfitting. Overfitting is the procedure through which model fits the parameters too closely with respect to the specific observations in the training set, does not generalize properly to new data, due to this reason the model is too complicated for the given training data. The primary goal of LDA is to locate the feature subspace that optimizes class separability. LDA is the more superior technique for classification tasks in comparison to the PCA [17].

Machine Learning Techniques used for Classification

We give a brief description of the Machine Learning techniques that we employed in the present work as following.

Neural Network

The Neural Network is one of the most important Machine Learning techniques for the purpose of classification of breast cancer into benign and malignant. Nowadays, Neural Network techniques are utilized for breast cancer diagnosis and prognosis. The Neural Network contains series of neurons. The Neural Networks are said to be inspired from the human brain, which utilizes, analyzes and processes various information. Neural Networks are made of neurons (perceptrons), which might be connected to each other. Perceptrons form the atomic unit of Neural Networks. They take a set of numbers as input, multiply the input with some weights and return an output. In the Neural Network, hidden neurons can be generated on the basis of some combination of the observed data. There are various activation functions such as Sigmoid function, Hyperbolic Tangent function, Radial Basis function and Gaussian function available in Neural Network [2, 19].

Deep Neural Network (DNN)

Deep Neural Network is a Feed Forward Neural Network with multiple hidden layers**.** A Deep Neural Network is capable to fit deep architecture more accurately with fewer parameters than Neural Network. One of the biggest challenges in training Deep Neural Network is how to efficiently learn the weights? Deep Neural Network is a very complex model [1, 19]. To train Deep Neural Networks with the help of R programming language, the same procedure was followed as what we already used to train Neural Networks.

Model Evaluation

The performance of the model can be evaluated using different parameters. The Confusion Matrix is used to quantify the overall performance of a model. However, there are several other performance measures that may be used to measure a model's relevance, such as Precision, recall, F1-score, Accuracy etc.

Confusion Matrix: Confusion Matrix is a matrix that lays out the performance of a Machine Learning technique. The Confusion Matrix is a square matrix that reports the counts of the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [14, 17].

Table 2 Confusion Matrix

	Test Outcomes	
Predictions	True Positives(TP) False Positives(FP)	False Negatives(FN) True Negatives(TN)

1. **True Positive (TP):** The prediction is actual.
2. **True Negative (TN):** No detection of abnormality in healthy man or woman.
3. **False Positive (FP):** Detection of abnormality is found in healthy man or woman.
4. **(iv)False Negative (FN):** No detection of malignant is found proves to be not true.

Next are the different measures, which refer to the model performance across all classes **[17, 19].**

Accuracy: Accuracy can be calculated as:
Accuracy=$(TP + TN)/(TN + TP + FP + FN)$

Specificity: Specificity can be calculated as:

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP})$$

Sensitivity: Sensitivity can be calculated as
Sensitivity=$TP/(TP + FN)$

Detection Rate: Percentage of True Positives.

Prediction Error (ERR): Prediction Error can be calculated as:

ERR = (FP+FN)/ (FP+FN+TP+TN)

False Positive Rate (FPR): False Positive Rate can be calculated as

False Positive Rate (FPR) = FP/ (FP+TN)

Precision (PRE): Precision (PRE) can be calculated as:

Precision=TP/ (TP+FP)

Recall: Recall is synonymous of Sensitivity**.** Recall (REC) can be calculated as:

Recall=Sensitivity= TP/ (FN+TP)

F1-score: F1-score can be calculated as:

F1-score=2((Precision*Recall)/ (Precision+ Recall))

Receiver Operating Characteristic(ROC): ROC graphs are useful tools for selecting models for classification based on their performance Based on the ROC Curve ,we can compute the ROC Area Under the Curve to characterize the performance of a classification model**[17].**

RESULT AND CONCLUSION

The dataset is randomly divided into two subsets (i) 70% of the instances to training (ii) 30% of instances to testing. The performance of Neural Network and Deep Neural Network techniques is evaluated using different statistical measures. In Table 3, the results are reported for Machine Learning techniques (Neural Network and Deep Neural Network). Table 4 exhibits the classification results of breast cancer into benign and malignant obtained by using confusion matrix of above Machine Learning techniques. Figure 1 depicts the graph that shows the comparison of Neural Network and Deep Neural Network (DNN) techniques in terms of different parameters. Figure 2, 3,4,5,6 and 7 showcases ROC graphs of Neural Network and Deep Neural Network techniques. By comparing classification results of Neural Network and Deep Neural Network, Neural Network technique without Dimensionality Reduction techniques gives better result in terms of Sensitivity or Recall (97.19%) and Detection Rate (35.29%). Neural Network with LDA gives better result in terms of Accuracy (97.06%), Specificity (100%), Precision (100%) and F1-Score (97.71%). Deep Neural Network without Dimensionality Reduction techniques also gives same result in terms of Specificity (100%) and Precision (100%). Deep Neural Network with LDA technique gives result in terms of Specificity (100%) and Prediction Error (37.05%). Deep Neural Network with PCA gives result in terms of FPR (12.30%).

Table 3

	Neural Network	Neural Network (PCA)	Neural Network (LDA)	Deep Neural Network	Deep Neural Network(LDA)	Deep Neural Network(PCA)
Accuracy (%)	96.47	95.29	97.06	96.34	62.94	91.76
Specificity (%)	95.23	95.08	100.00	100.0	00.00	87.69
Precision (%)	97.19	97.19	100.00	100.00	100.00	92.52
Recall (%)	97.19	95.41	95.53	94.39	62.94	94.28
F1 score (%)	97.19	96.29	97.71	97.11	77.25	93.39
Prediction Error (%)	03.52	04.70	02.94	03.65	37.05	08.23
FPR (%)	04.76	04.91	0.000	0.000	0.000	12.30
Detection Rate (%)	35.29	34.12	34.12	33.52	0.000	33.53

Table 4 Confusion Matrix for Machine Learning Techniques

Neural Network

	Test outcomes	
Predictions	Benign	Malignant
Benign	104(TP)	3(FN)
Malignant	3(FP)	60(TN)

Neural Network (PCA)

	Test outcomes	
Predictions	Benign	Malignant
Benign	104(TP)	5(FN)
Malignant	3(FP)	58(TN)

Neural Network (LDA)

	Test outcomes	
Predictions	Benign	Malignant
Benign	107(TP)	5(FN)
Malignant	0(FP)	58(TN)

DNN

	Test outcomes	
Predictions	Benign	Malignant
Benign	101(TP)	6(FN)
Malignant	0(FP)	57(TN)

DNN (PCA)

	Test outcomes	
Predictions	Benign	Malignant
Benign	99(TP)	6(FN)
Malignant	8(FP)	57(TN)

DNN (LDA)

	Test outcomes	
Predictions	Benign	Malignant
Benign	107(TP)	63(FN)
Malignant	0(FP)	0(TN)

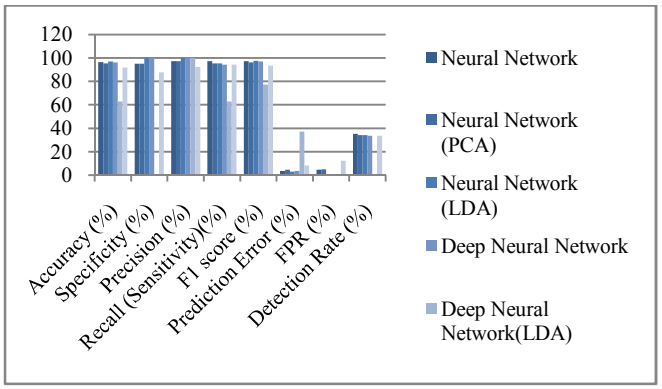

Figure 1

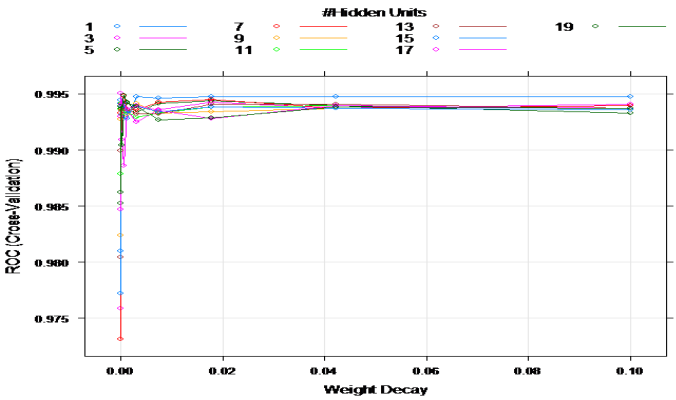

Figure 2 Neural Network

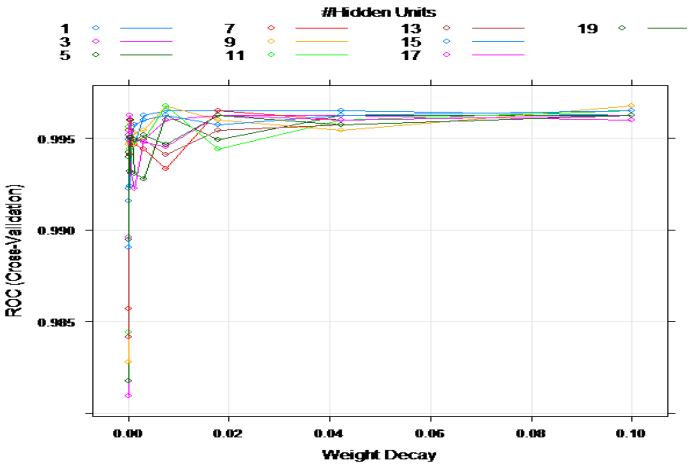

Figure 3 Neural Network (PCA)

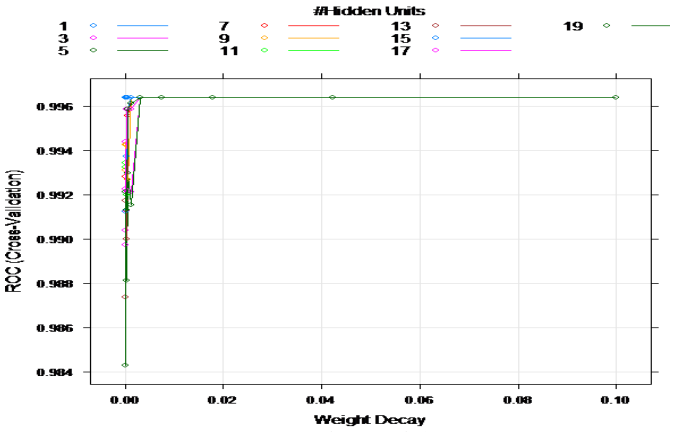

Figure 4 Neural Network (LDA)

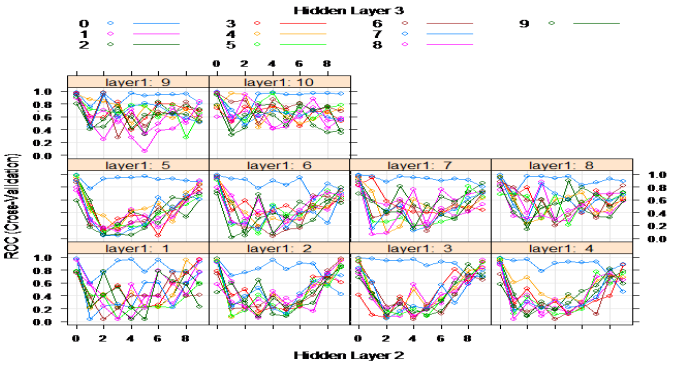

Figure 5 Deep Neural Network

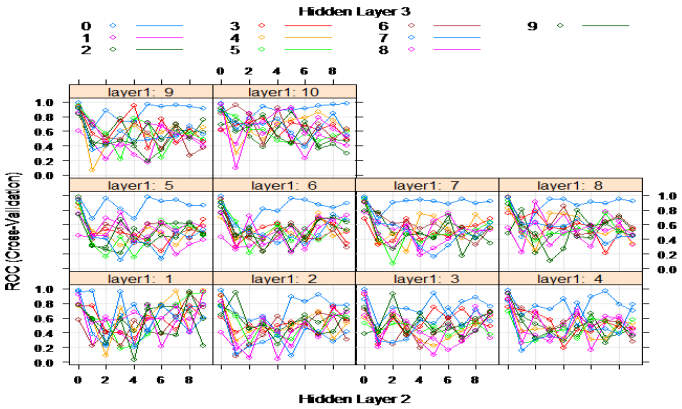

Figure 6 Deep Neural Network (PCA)

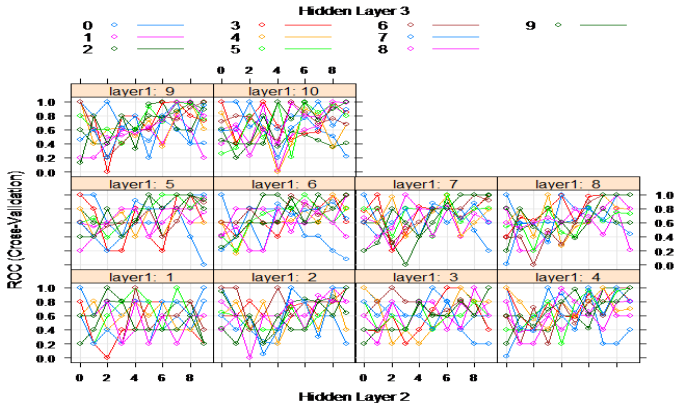

Figure 7 Deep Neural Network (LDA)

CONCLUSION

The present study compares the performance measures of two machine learning techniques: Neural Network and Deep Neural Network on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Neural Network and Deep Neural Network both techniques exhibited high performance on the classification of breast cancer in terms of different parameters. In future work, we propose to analyze K-Nearest Neighbor (KNN) technique with and without dimensionality reduction techniques.

References

1. Ahmed M. Abdel Zaher and Ayman M. Eldeib (2015),"Breast Cancer Classification Using Deep Belief Networks", Expert Systems with Applications.
2. Alias Anu, Paulchamy B. (2014),"Detection of Breast Cancer Using Artificial Neural Networks", IJIRSET, Vol.3, Issue 3.
3. Baoyu Zheng, Wei Qian, and Laurence P. Clarke (1996),"Digital Mammography: Mixed Feature Neural Network with Spectral Entropy Decision for Detection of Microcalcifications", IEEE Transactions on medical imaging, Volume 15, No 5,Page(s)-589-597.
4. Beg Mirza Mahjabeen and Jain Monika**,** **"**An Analysis of the Methods Employed for Breast Cancer Diagnosis", Page(s)-1-5.
5. Breast Cancer Wisconsin Data[online]. Available: http://archive.ics.uci.edu/ml/datsets/Breast+Cancer+ Wisconsin+ (Diagnostic).
6. Chen Ren Dar, Chang Feng Ruey and Huang Len Yu (2000),"Breast Cancer Diagnosis using Self Organizing Map for Sonography", Elsevier Ultrasound in Medical and Biology, Volume-26, Page(s)-405-411.
7. Cheng D.H., Shan Juan, Ju Wen, Guo Yanhui and Zhang Ling (2010),"Automated Breast Cancer Detection and Classification Using Ultrasound Images: A survey", Science Direct, Page(s)-299-317.
8. Endo A., Shibata T. and Tanaka H. (2008), "Comparison of seven algorithms to predict breast cancer survival", Biomedical Soft Computing and Human Sciences, vol.13, Page(s)-11-16.
9. https://www.cancer.org.
10. https://www.medicinenet.com/article/Breast Cancer.
11. https://www.niehs.nih.gov/Breast Cancer, "National Institute of Environmental Health Sciences".
12. Jerry R.Balentine (2018),"Breast Cancer Causes, Types, Symptoms, Signs, Stages, Treatment". WebMD Newsletters.
13. Jihong Liu and Weina Ma (2007),"An Effective Recognition Method of Breast Cancer Based on PCA and SVM Algorithm", Springer ICMB, Page(s) - 57-64.
14. Joshi Arpita and Mehta Ashish (2017),"Comparative Analysis of Various Machine Learning Techniques for Diagnosis of Breast Cancer", *International Journal on Emerging Technologies"*, Page(s)-522-526.
15. Karabatak Murat and Ince Cevdet M.(2009),"An Expert System for Detection of Breast Cancer based on Association Rules and Neural Network", ELSEVIER,ScienceDirect,Page 3465-3469.
16. Male Breast Cancer-National Breast Cancer Foundation, www.nationalbreastcancer.org.
17. Raschka Sebastian and Mirjalili Vahid (2017),"Python Machine Learning", 2nd Edition, Packt Publishing.ISBN:978-1-78712-593-3.
18. Sharma N.Ganesh, Dave Rahul, Sanadya Jyotsana, Sharma Piush and Sharma K.K. (2010),"Various Types and Management of Breast Cancer: An Overview", *Journal of Advanced Pharmaceutical Technology and Research*, Page(s)-109-126.
19. Wiley F. Joshua (2016),"R Deep Learning Essentials", Packt Publishing.ISBN: 978-1-78528-058-0.
20. Yan Xu,Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai and Eric I-Chao Chang(2014),"Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis", IEEE International Conference on Acoustic, Speech and Signal Processing.

How to cite this article:

Arpita Joshi and Ashish Mehta.2018, Breast Cancer Data Classification Using Neural Network and Deep Neural Network Techniques. *Int J Recent Sci Res.* 9(4), pp. 25788-25792. DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.1931

****