International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 5(1), September - October 2015 ISSN 2278-6856

Diagnosis of Breast Cancer using SOM Neural Network

[1]Praveen C. Shetiye, [2]Ashok A. Ghatol , [3]Vilas N. Ghate

[1]Govt .College of Engineering Karad Maharashtra, India

[2]former VC of Babasaheb Technical University Lonere Maharashtra India

[3]Govt .College of Engineering Chandrapur Maharashtra, India

Abstract

A neural network system is designed and optimize to analyze the quantitative data from the impedance spectrum from where the breast tissue features are computed. These data was used to predict and classify the breast cancer Car (carcinoma), fad (fibro-adenoma+ mastopathy + glandular), Con (connective), Adi (adipose). The performance of an artificial neural network (ANN) is verified with nine quantitative parameters computed from Impedance measurements. The Self Organizing Feature Map (SOM) network is trained using the different data partitioning methods and tested its performance on seen and unseen data in terms of classification accuracy, MSE and correlation coefficient. The network is yielded better classification accuracy (93.75%) with testing and cross validation MSE of 0.00044 and 0.00025 respectively.

Index Terms:-Breast cancer, Electrical impedance, SOM

1.INTRODUCTION

Breast cancer is the second most common disease that affects women, next to skin cancer. However, it is curable at initial stages and hence early detection can reduce the mortality rate. Conventionally, Breast cancer detection and classification is performed by a clinician or a pathologist by observing stained biopsy images under the microscope. However, this is time consuming and can lead to erroneous results [1]. Some evidence has been found that malignant breast tumors have lower electrical impedance than surrounding normal tissues. Although the separation of malignant tumors from benign lesions based on impedance measurements needs further investigation, electrical impedance could be used as an indicator for breast cancer detection. Biological tissues have complex electrical impedance related to the tissue dimension, the internal structure and the arrangement of the constituent cells. Therefore, the electrical impedance can provide useful information based on heterogeneous tissue structures, physiological states and functions [2, 3]. In addition the concepts of time varying distribution of electrical properties inside a human body such as electrical conductivity and (or) permittivity can be used to analyze a variety of medical conditions. High-conductivity materials allow the passage of both direct and alternating currents and high-permittivity materials allow the passage of only alternating currents. Both of these properties are of interest in medical systems since different tissues have different conductivities and permittivities [4-13].

2.NEURAL NETWORK BASED CLASSIFIER

During the recent times, because of their discriminative training ability and easy implementation, the Artificial Neural Networks (ANN) find extensive use in classification of the type of tumor in breast cancer problem. It turns out that the selection of number of nodes for an ANN is an important criterion in breast cancer analysis. However, a large network means more computational expenses, resulting in more hardware and time related cost. Therefore a compact and optimum design of neural network is needed towards real time detection of tumor type in breast cancer analysis.[14-18]

1.1 Data Collection

For this research work public available dataset is used [19]. Impedance measurements of freshly excised breast tissue were made at the following frequencies: 15.625, 31.25, 62.5, 125, 250, 500, 1000 KHz. These measurements plotted in the (real, -imaginary) plane constitute the impedance spectrum from where the breast tissue features are computed. The dataset can be used for predicting the classification of either the original 6 classes or of 4 classes by merging together the fibro-adenoma, mastopathy and glandular classes whose discrimination is not important (they cannot be accurately discriminated anyway). The following features are taken for the classification of four classes,

I_0 - Impedivity (ohm) at zero frequency
P_A- Phase angle at 500 KHz
HFS - High-frequency slope of phase angle
DA - Impedance distance between spectral ends
A- Area under spectrum
A/DA - Area normalized by DA
MAX IP - Maximum of the spectrum
DR- Distance between I_0 and real part of the maximum frequency point
P - Length of the spectral curve

Total Data is to be classified into four class as, Car (carcinoma), fad (fibro-adenoma+ mastopathy + glandular), Con (connective), Adi (adipose).

1.2 Design and Optimization of SOM Neural Network Classifier

The Kohonen self-organizing map (SOM) network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input. This means that points close to each other in the input space are mapped to the same or neighboring PEs in the output space. The basis of the Kohonen SOM network is soft competition among the PEs in the output space. The Kohonen SOM is a fully connected, single-layer linear network. The output generally is organized in a one or two-dimensional arrangement of PEs, which are called neighborhoods. All the units in the neighborhood that receive positive feedback from the winning unit participate in the learning process. Even if a neighboring unit's weight is orthogonal to the input vector, its weight vector will still change in response to the input vector. This simple addition to the competitive process is sufficient to account for the order mapping. The general learning algorithm used is as follows,

Step 1: Set the topological neighborhood parameters. Set learning rate and initialize weights.
Step 2: While stopping condition is false, do steps 3-9
Step 3: For each input vector X do steps 4-6
Step 4: For each *j* compute squared Euclidean distance.

$$D(j) = \sum (w_{ij} - x_i)^2 \qquad (1)$$

$i = 1$ to n and $j = 1$ to m
Step 5: Find index *J* when *D(j)* is minimum.
Step 6: For all units J with the specified neighborhood of J and for all I update the weights.

$$w_{ij(new)} = w_{ij(old)} + \alpha\left[x_i - w_{ij(old)}\right] \qquad (2)$$

Step 7: Update the learning rate
Step 8: Reduce the radius of topological neighborhood at specified times.
Step 9: Test the stopping condition.

Selection of error criterion:
Supervised learning requires a metric, a measure of how the network is doing. Members of the Error Criteria family monitor the output of a network, compare it with some desired response and report any error to the appropriate learning procedure. In gradient descent learning, the metric is determined by calculating the sensitivity that a cost function has with respect to the network's output. This cost function, *J*, is normally positive, but should decay towards zero as the network approaches the desired response. The literature has presented several cost functions, in which *p* is to be define such as p=1, 2, 3, 4… ∞ criterion is L_1, L_2, L_3, L_4 …L_∞ Components in the Error Criteria family are defined by a cost function of the form:

$$J(t) = \frac{1}{2}\sum_{i=1}(d_i(t) - y_i(t))^p \qquad (3)$$

and error function:

$$e_i(t) = (d_i(t) - y_i(t)) \qquad (4)$$

Where *d(t)* and *y(t)* are the desired response and network's output, respectively. To select the correct error criterion various error criterion has been tested and results are shown in Fig. 1. Hence L_2 error criterion is selected.

To classify the four conditions, four PEs are in the output layer. Remaining parametrs of network, such as Neighborhood shape, number of rows and columns of Neighborhood shape, number of processing elements in the hidden layer, step size and momentum of hidden and output layer are selected and optimised by expermentations and the results are as shown in Fig.2 to Fig.7.

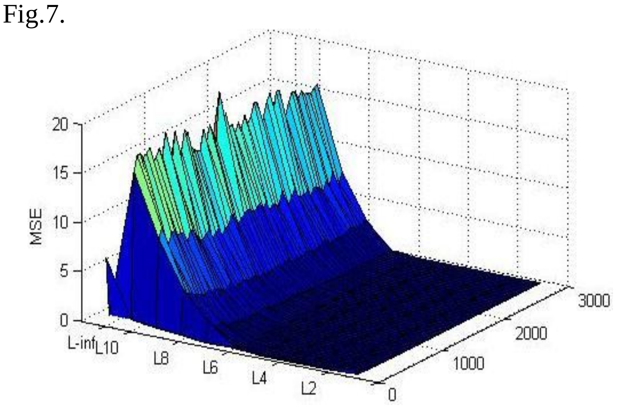

Fig.1. Variation of Average MSE with Error Criterion

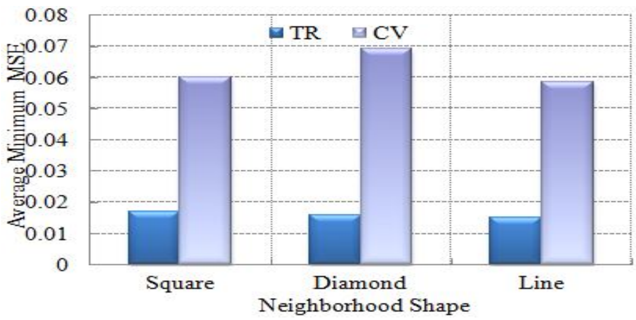

Fig.2. Variation of Average minimum MSE with Neighborhood shape

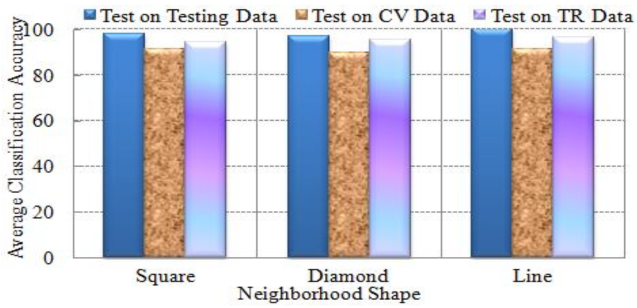

Fig.3.Variation of Average classification accuracy with Neighborhood shape

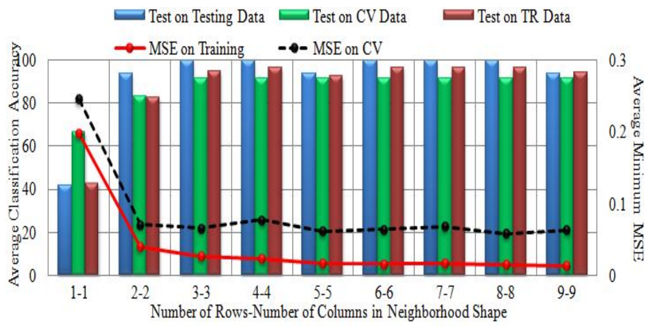

Fig.4. Variation of Average minimum MSE and classification Accuracy with number of rows and columns of Neighborhood

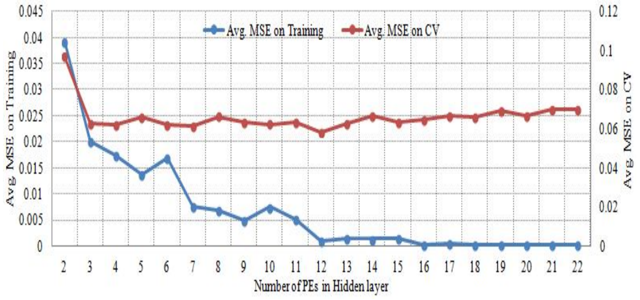

Fig.5. Variation of average MSE with number of processing elements in the hidden layer

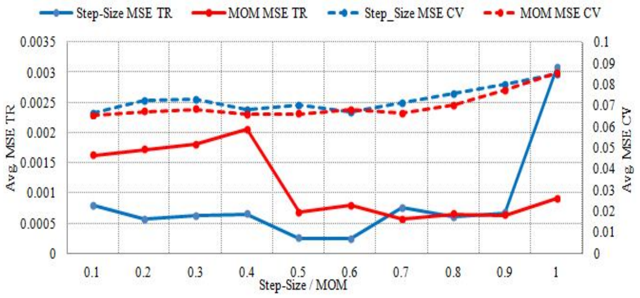

Fig.6. Variation of Average Minimum MSE with step size and momentum rate of Output layer and output layer

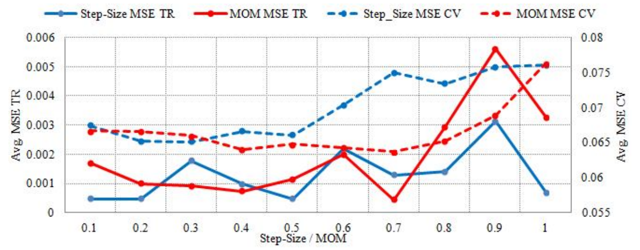

Fig.7. Variation of Average Minimum MSE with step size and momentum rate of Hidden layer and output layer

From the results it is observed that Line neighborhood shape with 8 rows and 8 columns and starting radius 7 and final radius 0 gives the optimum results. For supervised learning the network is retrain for five times with different number hidden layers and number of PEs in hidden layers. It is observed that single hidden layer with 12 PEs in hidden layer gives better results. Similar experimentations are performed to decide the step size and momentum rate of hidden layer and output layer considering the average minimum MSE as performance index.

Finally with above experimentations SOM-NN is designed with following specifications,
Neighborhood Shape: Line,
Number of Rows in Neighborhood Shape: 8,
Number of columns in Neighborhood Shape: 8,
Starting Radius of Neighborhood Shape: 7,
Final Radius of Neighborhood Shape: 0,
Number of epochs = 10000,
Exemplars for training = 70%,
Exemplars for cross validation = 15%,
Exemplars for Testing = 15%
Number of connection weights: 449
Number of PEs in Hidden Layer: 12
Time elapsed per epoch per exemplar: 0.716 ms
Hidden layer:
T.F. :Tanh , Learning Rule : Momentum
Stepsize: 0.5 Momentum : 0.7
Output layer:
T.F. :Tanh , Learning Rule : Momentum
Stepsize: 0.6 Momentum: 0.7

Different datasets are formed using variable split ratios and leave-N-out cross validation technique. Proposed NN is trained on various datasets and later validated carefully so as to ensure that its performance does not depend on specific data partitioning scheme. The performance of the NN should be consistently optimal over all the datasets with respect to MSE and classification accuracy. Finally designed SOM NN is trained five times with different random weight initialization and tested on Testing dataset, CV dataset and Training dataset. For training and testing the Leave-N-Out method, data tagging by percent and data tagging by various groups are used. Leave-N-Out training is a technique that allows one to evaluate how well the model generalizes. It also is very useful for small data sets, since it allows one to use the entire data set for training and testing. The algorithm trains the network multiple times, each time omitting a different subset of the data and using that subset for testing. The outputs from each tested subset are combined into one testing report and the model is trained one additional time using all of the data. The set of weights saved from the final training run can then be used for additional testing. To check the learning ability and classification accuracy the total data is divided in four groups. First two groups (50% data) are tagged as Training data and third and forth group (each 25%) is tagged for Cross Validation and Testing (1234:1,2-TR, 3-CV, 4-Test). Similar 24 combinations are prepared and network is train and test for each group. Results are shown in Fig.8 and Fig.9

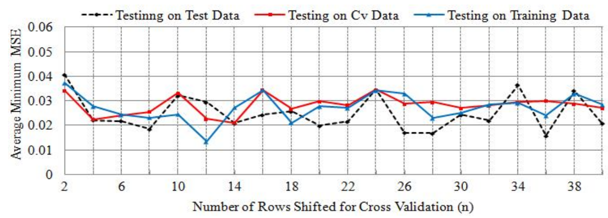

Fig.8. Variation of average MSE with Test on Testing, CV and Training dataset with CV rows shifted (n)

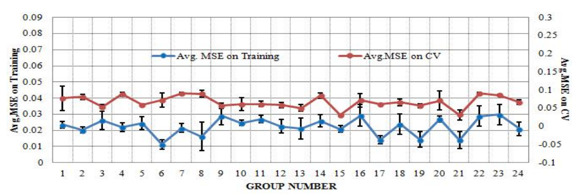

Fig.9. Variation of average Minimum MSE with Training and CV with group of Dataset

3.RESULTS AND DISCUSSION

In this paper author examines the results of SOM NN based classifier. The proposed SOM NN classifier, all samples including cross validation and testing samples are correctly classified with good classification accuracy to the corresponding classes (classification accuracy is 93.73% on testing and cross validation data). Furthermore, observing the classification errors of the two approaches, it can be seen that the classification error of the proposed SOM NN is 0.00044 and 0.00025 on testing and CV respectively, which is reasonably good. These results show that the proposed SOM NN can reliably recognize class of breast cancer. The comparison results prove that the proposed SOM NN have obtained significant achievements in recognition accuracy and provided a better generalization capability.

References

[1] Shekhar Singh, Dr P. R. Gupta, “Breast Cancer detection and Classification using Neural Network”, International Journal Of Advanced Engineering Sciences And Technologies, Vol No. 6, Issue No. 1, pp 4 – 9.

[2] Walker D. C. Et al, “Modelling electrical impedivity of normal and premaligant cervical tissue”, Electronic Letters, 36(19):1603-1604, 2000

[3] Walker D. C. Et al, “Modelled current distribution in cervical squamous tissue”, Physiological Measurement, 23(1): 159-168, 2002

[4] M. Cheney, D. Isaacson, and J. C. Newell, “Electrical impedance tomography,” SIAM Review, 41(1): 85–101, 1999.

[5] S Kimura, T Morimoto, T Uyama, Y Monden, Y Kinouchi and T Iritani, “Application of electrical impedance analysis for diagnosis of a pulmonary mass”, Chest Journal, 105: 16791682, 1995

[6] Y. Kinouchi, T. Iritani, T. Morimoto et al, “Fast in vivo measurement of local tissue impedance using needle electrode”, Medical and Biological Engineering and Computing, 35: 486-492, 1997

[7] Xueli Zhao, Y. Kinouchi, E. Yasuno, “A new method for noninvasive measurement of multilayer tissue conductivity and structure using divided electrode”, IEEE Transaction on Biomedical Engineering, 51(2): 362-370, 2004

[8] H. Kato, E. Yasuno, Y. Kinouchi, “Electrical impedance tomography for local bological tissue”, Proceeding of 8th International Conference on Control Robotics, Automation and Vision, 942946, 2004

[9] Malich A, Fritsch T, Anderson R, Boehm T, Freesmeyer MG, Fleck M, Kaiser WA: Electrical impedance scanning for classifying suspicious lesions: first results. Eur Radiol 2000, 10:1555-1561.

[10] T. Tang, S. U. Zhang, R. J. Sadleir, “A portable 8-electrode EIT measurement system”, IFMBE Proceeding, 3980-3983, 2007

[11] W. Yan, S. Hong, Z. Shu, R. Chaoshi, “Influences of compound electrode parameter on measurement sensitivity and reconstruction quality in electrical impedance tomography”, Proceedings of International Federation for Medical and Biological Engineering, 6, 2007

[12] A.V. Korjenevsky, "Electrical impedance tomography: research, medical applications and commercialization", Troitsk Conference on Medical, Physics and Innovations in Medicine", 2006

[13] B. Blad, P. Wendel, M. Jonsson, and K. Lindstrom, “An electrical impedance index to distinguish normal and cancerous tissues,” J. Med. Eng. Technol., vol. 2, no. 23, pp. 57–62, 1999.

[14] Jossinet J, Schmitt M: A review of parameters for the bioelectrical characterization of breast tissue. Ann NY Acad Sci 1999, 873:30-41

[15] Piperno G, Frei EH, Moshitzky M: Breast cancer screening by impedance measurements. Frontiers Med Biol Eng 1990, 2:111-117

[16] S. Ollmar, I. Nicander, J. Ollmar, and L. Emtestam, Information in full and reduced data sets of electrical impedance spectra from various skin conditions using a holographic neural network, Med. Biol. Eng. Comput., vol. 35, pp. 415419, 1997.

[17] Martin G, Martin R, Brieva MJ, Santamaria L: Electrical impedance scanning in breast cancer imaging: correlation with mammographic and histologic diagnosis. Eur Radiol 2002, 12:1471-1478.

[18] Laurence Fauseet, “Fundamentals of Neural Network, 2nd Edition” Prentice Hall. 1998.

[19] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

AUTHOR

Dr.V.N.Ghate received the B.E. degree in electrical Engineering from Sant Gadge Baba Amravati University, Amravati, India, in 1990, the M.E. degree in Control Systems from Shivaji University, Kolhapur, India, in 1993, and the Ph.D. degree from Sant Gadge Baba Amravati University in 2010. From 1994 to 2001, he was with the Department of Electrical Engineering, Government Polytechnic, Technical Education Government of Maharashtra. Since 2002, he has been with Department of Electrical Engineering, Government College of Engineering, Technical Education Government of Maharashtra. His areas of interest include neural network, and Electrical Machines and Drives

Dr. Ghate is Fellow of Institution of Electronics and Telecommunication Engineers (India), Institution of Engineers (India) and member of Instrument Society of India, IEEE and the Indian Society for Technical Education (India).

P.C. Shetiye received the B.E. degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, India, in 1990, M.B.A. degree from Nagpur University, Nagpur, India, 1997, M.E. degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, India, in 2001, pursing Ph.D. degree from Sant

Gadge Baba Amravati University. From 1993 to 1995, he was with department of Computer Science and Engineering, Government College of Engineering, Amravati, from 1996 to 2001, he was with the Department of Computer Science and Engineering, Government Polytechnic, Amravati, later on worked with MSBTE, Technical Education Government of Maharashtra. Since 2002, he has been with Department of Computer Science and Engineering [MCA], Government College of Engineering, Karad, Technical Education Government of Maharashtra. His areas of interest include AI, neural network, and System software.

Dr. A.A. Ghatol born on 29th August, 1948 is presently Vice-Chancellor of Dr. Babasaheb Ambedkar Technological University, Lonere-Raigad. Before joining as Vice-Chancellor, he was Principal/Director at College of Engineering, Pune during 2001-2005 and Principal at Government College of Engineering, Amaravati during 1994-2001. He holds a B.E. (Electrical), Nagpur University, M.Tech. (Electrical) and Ph.D Electrical from IIT, Mumbai in the field of High Power Semiconductor Devices. Over the last 33 years, he has been actively involved in the field of Technical Education as Academician, Researcher, Teacher, Planner and Administrator and handled various aspects of Technical Education including Management, Engineering, Technology etc. Best Teacher Award, of Govt. of Maharashtra was bestowed upon him during the academic year 1998 – 99. He has also received Quality Environment Award in the year 2002. He has to his credit 12 Ph.D. Students, 3 books and number of research scholars working under him for Ph.D. degree. Under his able and effective leadership, the University reached to a newer height and gained a feather in the cap by receiving the UGC recognition under section 12(B). Taking into consideration his active involvement in technical education, AICTE also entrusted a responsibility to Act as Chairman of Western Regional Council for three states for advising and assisting the council on issues and strategies on quality and quantity in the technical education. He is also the Executive Member of Indian Society of Technical Education.