Detection of Breast Cancer Using Electrical Impedance and RBF Neural Network

P. C. Shetiye, A. A. Ghatol, V. N. Ghate, and S. R. Patil

Abstract—**Breast cancer presents a serious medical and social problem worldwide. Early detection is a key to effective breast cancer treatment. Electrical Impedance Tomography (EIT) is a medical imaging technique that reconstructs internal electrical conductivity distribution of a body from impedance data, which can be used for detection of breast cancer. In this paper publicly available data of breast cancer is used to design and optimized Radial Basis Function (RBF) Neural Network based Classifier. Classifier is tested for generalization, which proves as an elegant classifier with the accuracy of 93%.**

Index Terms—**Breast cancer, electrical impedance, RBF.**

I. Introduction

Breast cancer presents a serious medical and social problem worldwide. Early detection is key to effective breast cancer treatment. Therefore, scientists are consistently looking for new diagnostic techniques that would be more efficient, easy to use and safe for the patient. Studies have shown that early diagnosis by means of breast imaging, including digital mammography, ultrasound imaging, and magnetic resonance imaging (MRI), could help prognosis and increase therapeutic options [1]. Relatively new technique of Electrical Impedance Tomography (EIT) has attracted much interest as low cost, non-invasive imaging tool for early diagnosis of breast cancer.

Electrical Impedance Tomography (EIT) is a medical imaging technique that reconstructs internal electrical conductivity distribution of a body from impedance data that is measured on the body surface, and Electrical Impedance Mammography (EIM) is the technique that applies EIT in breast cancer detection. EIM has the potential in detecting early stage cancer; however there are still challenges that hindering EIM to be provided as a routine health care system. Electrical Impedance refers to the measurement in real and imaginary units [2]-[4].

The most widely adopted EIT measurement method utilizes electrodes on the object surface as a source of injecting small alternating current (AC) and as a voltage recorder, which measures the differences in electric potential. Apart from measuring signal magnitude, the phase angle difference between the injecting current and recorded voltage may also be captured as the impedance. 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 [5]-[7].

II. Neural Network Based Classifier

Recently, artificial neural networks have been applied to classifying mammographic masses for early-stage breast cancer detection and diagnosis [8]-[10] which would help reduce the number of unnecessary surgical biopsies. Artificial neural networks, with the properties of experience-based learning and generalization ability, are regarded as one of the emerging computational technologies for solving complex problems that might not have a tractable solution provided by traditional methods [11], [12].

A. Data Collection

For this research work public available dataset is used [13]. 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,

I0 - Impedivity (ohm) at zero frequency.

PA - 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 I0 and real part of the max 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).

B. Design and Optimization of RBF Neural Network Classifier

Radial basis function (RBF) networks are nonlinear hybrid networks typically containing a single hidden layer of processing elements (PEs). This layer uses gaussian transfer

Manuscript received October 8, 2014; revised December 26, 2014.

P. C. Shetiye, is with Government College of Engineering Karad, Maharashtra, 415124, India (e-mail: praveen.shetiye@gmail.com).

A. A. Ghatol was with Dr. B. A. T. U. Lonere, India (e-mail: aaghatol@gmail.com).

V. N. Ghate and S. R. Patil are with Government College of Engineering Chandrapur, Maharashtra, 442403, India (e-mail: vng786@gmail.com, sanjaypatil_71@yahoo.com).

functions, rather than the standard sigmoidal functions employed by MLPs. The centers and widths of the gaussians are set by unsupervised learning rules, and supervised learning is applied to the output layer. These networks tend to learn much faster than MLPs. Use of this network is recommended only when the number of exemplars so dispersed that clustering is ill-defined. For standard RBF's, the supervised segment of the network only needs to produce a linear combination of the output at the unsupervised layer. The selections for the learning address two peculiarities of competitive learning. Competitive learning has an intrinsic metric. Competitive learning keeps an intrinsic probability distribution of the input data. It has the drawback that some PEs may never fire, while others may always win the competition. To avoid these extremes, one can include a conscience mechanism that keeps a count on how often a PE wins the competition, and enforces a constant wining rate across the PEs. The centers of the Gaussians are placed with a conscience mechanism. The general learning algorithm is as follows:

The response of Gaussian activation function is nonnegative for all value of *x*. The function is defined as

$$f(x)=\exp(-x^2) \tag{1}$$

and its derivative

$$f'(x)=-2x\exp(-x^2)=-2f(x) \tag{2}$$

The radial basis function is different from the back propagation network in the Gaussian function it uses.

Step 1: Initialize the weights (set to small values)

Step 2: While stopping condition is false do step3-10

Step 3: for each input do step 4-9

Step 4: Each input unit (x_i, i=1,…, n) receives input signals to all units in the hidden layer.

Step 5: Calculate the radial basis function

Step 6: Choose the centers for the radial basis functions. The centers are chosen from the set of input vectors. A sufficient numbers of centers have been selected in order to ensure adequate sampling of the input vector space.

Step 7: The output of i_m unit $v_i(x_i)$ in the hidden layer

$$v_i(x_i)=e\left\{\sum_{j=1}^{r}\frac{\left[x_{ji}-(x'_{ji})^2\right]}{\sigma_i^2}\right\} \tag{3}$$

where

x_{ji} -center of the RBF unit for input variables

x'_{ji} - j^{th} variable of input pattern

σ_i -Width of the i^{th} RBF unit

Step 8: Initialize the weights in the output layer of the network to some small random value

Step 9: Calculate the output of the neural network

$$y_{net}=\sum_{i=1}^{H}w_{im}v_i(x_i)+w_0 \tag{4}$$

where

H-number of hidden layer nodes (RBF Function)

y_{net} -Output value of m^{th} node in output layer for the n^{th} incoming pattern

w_{im} - Weight between i^{th} RBF unit and m^{th} output node

w_0 - Biasing term at n^{th} output node

Step 10: Calculate the error and check the stopping condition

The randomized data is fed to the neural network and is retrained five times with different random weight initialization so as to remove biasing and ensure true learning and generalization .From these experimentations the unsupervised and supervised parameters, competitive rule, metric and cluster centers have been selected. Experimentation results are shown in Fig. 1 to Fig. 7.

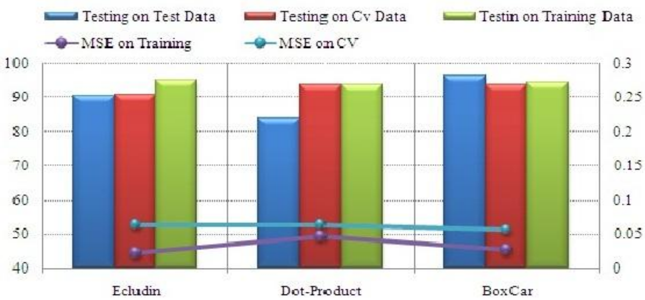

Fig. 1. Variation of average classification accuracy and MSE with testing on testing, CV and training dataset and metric with cons. full as competitive rule.

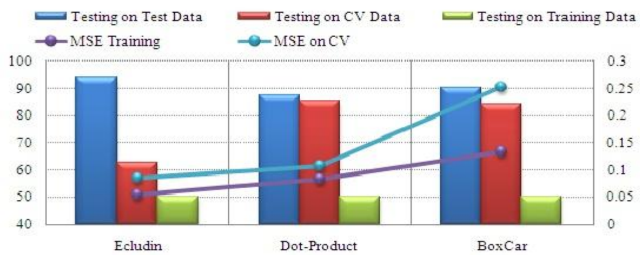

Fig. 2. Variation of average classification accuracy and MSE with testing on testing, CV and training dataset and metric with STD. full as competitive rule.

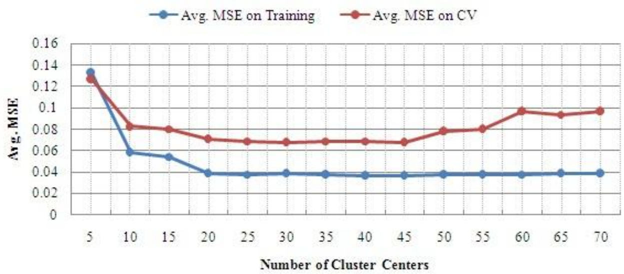

Fig. 3. Variation of average minimum MSE with number of cluster centers.

Fig. 4. Variation of average minimum MSE on training and CV dataset with step size of output layer.

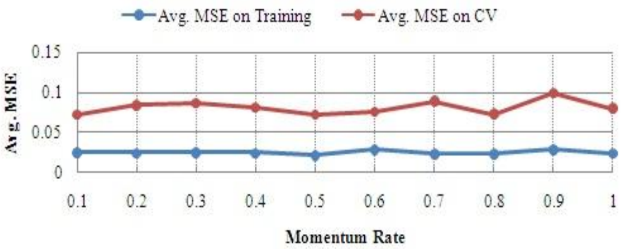

Fig. 5. Variation of average minimum MSE on training and CV dataset with momentum of output layer.

Fig. 6. Variation of average classification accuracy and MSE with testing on testing data set with percent data tagged for training.

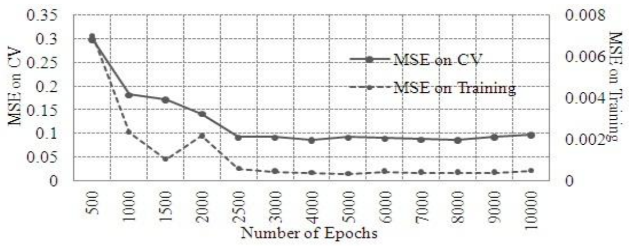

Fig. 7. Variation of average minimum MSE on training and CV dataset with number of cluster centers.

From the above experimentation results it is seen that BoxCar Metric with Cons. Full as Competitive Rule gives the minimum MSE and maximum classification accuracy. Variation of Average Minimum MSE with number of Cluster Centers shows that after 45 cluster centers there is no significant change in training as well as CV MSE but time required for training with higher number of cluster centers will reasonably higher. Finally, the optimized RBF classifier is designed with the following parameters:

Number of Input PEs: 9
Number of Output PEs: 4
Competitive Rule: Conscience Full,
Metric: Boxcar,
Output Layer: Learning Rule: Momentum, Transfer
Function: Tanh Step Size: 0 .600, Momentum: 0.5000
Unsupervised Learning: Maximum Epoch: 1000
Learning Rate: start at: 0.01 **Decay to:** 0.001
Number of Cluster Centers: 45
Number of epochs: 4000
Exemplars for training: 70%,
Exemplars for cross validation : 15%
Exemplars for Testing : 15%
Number of connection weights: 634

III. Generalization of Optimized NN Based Classifier

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 RBF 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 to Fig. 11

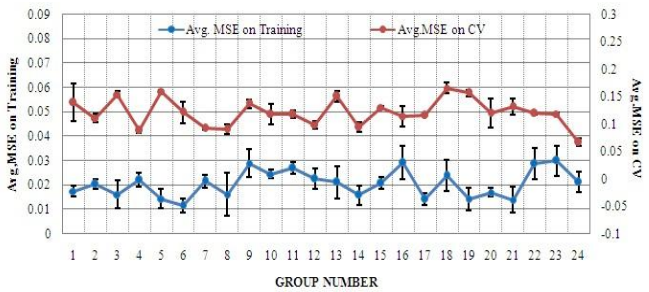

Fig. 8. Variation of average minimum MSE with training and CV with group of dataset.

Fig. 9. Variation of average classification accuracy with training and CV with group of dataset.

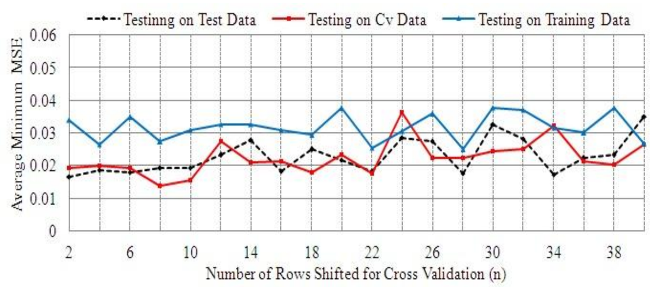

Fig. 10. Variation of average minimum MSE with test on testing, CV and training dataset with CV rows shifted (*n*).

Experimentations on generalization of classifier shows the consistency in average minimum MSE and classification accuracy on testing on Test Data, CV data and Training Data for different sequence of data presented, which indicate the designed classifier is learned and not memorized.

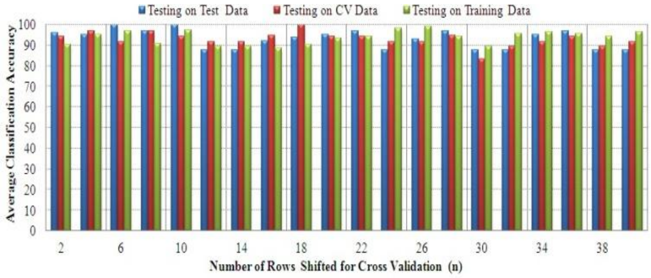

Fig. 11. Variation of average classification accuracy with test on testing, CV and training dataset with CV rows shifted (*n*).

IV. Conclusion

In this paper, the authors evaluated the performance of the developed RBF NN based classifier for detection of four classes of breast cancer and examined the results. Experimentations are performed on RBF neural network model and for this network, combination of Conscience Full Competitive Rule and Boxcar metric gives the best results. Initially, MSE goes on decreasing with increase in number of cluster centers and it attains minimum value at 45 cluster centers, beyond which it is observed that there is no significance decrease in MSE. For stopping the training of the network the number of epochs, is one of the early stopping criteria used and since no significant change in MSE 4000 epochs are selected for the training.

From the analysis, it is seen that RBF NN works as an elegant classifier for detection of four classes of breast cancer, in the sense that, average MSE on testing and cross validation samples is consistently observed as reasonably low such as 0.022941 and 0.0224, respectively. In addition, average classification accuracy on testing as well as cross validation instances is obtained as 93.07% and 92.82%, respectively indicating a reasonable classification.

References

[1] W. L. Donegan, J. S. Spratt, and A. Orsini, *Cancer of the Breast*, 5th ed., Amsterdam, Netherlands: Elsevier Science, 2002.

[2] *The Biomedical Engineering Handbook*, Second Edition, vol. 68, CRC Press, 2000, pp. 68-1-68-13.

[3] D. C. Walker *et al.*, "Modelling electrical impedivity of normal and premaligant cervical tissue," *Electronic Letters*, vol. 36, no. 19, pp. 1603-1604, 2000.

[4] D. C. Walker *et al.*, "Modelled current distribution in cervical squamous tissue," *Physiological Measurement*, vol. 23, no. 1, pp. 159-168, 2002.

[5] J. Campbell and N. Dimache, "3D EIT - MEIK in clinical application: Observations and preliminary results," in *Proc. 2006 World Congress on Medical Physics and Biomedical Engineering*, pp. 3906-3910, vol. 14, 2007.

[6] R. J. Halter, A. Hartov, and K. D. Paulsen, "A broadband high-frequency electrical impedance tomography system for breast imaging," *IEEE Trans Biomed Eng*, vol. 55, no. 2, pp. 650-659, 2008.

[7] N. Huber *et al.*, "Further investigation of a contactless patient-electrode interface of an electrical impedance mammography system," presented at the 10th International Conference on Biomedical Applications of Electrical Impedance Tomography (EIT 2009), School of Mathematics, The University of Manchester, Manchester, UK, June 15-19, 2009.

[8] 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.

[9] J. Jossinet and M. Schmitt, "A review of parameters for the bioelectrical characterization of breast tissue," *Ann NY Acad Sci.*, vol. 873, pp. 30-41, Apr. 1999.

[10] G. Piperno, E. H. Frei, and M. Moshitzky, "Breast cancer screening by impedance measurements," *Frontiers Med Biol Eng*, vol. 2, no. 2, pp. 111-117, 1990.

[11] 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," *J. Med. Biol. Eng. Comput.*, vol. 35, no. 4, pp. 415-419, July 1997.

[12] G. Martin, R. Martin, M. J. Brieva, and L. Santamaria, "Electrical impedance scanning in breast cancer imaging: Correlation with mammographic and histologic diagnosis," *Eur Radiol*, vol. 12, no. 6, pp. 1471-1478, 2002.

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

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 the 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 a 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.

Sanjay R. Patil was born in Sangli, India in 1971. He received the B.E. degree in instrumentation engineering from Shivaji University, Kolhapur, in 1992 and the M.Tech. degree in control and guidance from Indian Institute of Technology, Roorkee, India, in 2009. From 1992 to 1994 he was with SNJP Polytechnic, Karnataka. From 1994 to 1997 he worked as a senior instrumentation engineer at Apex Industries and from 1997 to 2002 he was faculty at Dr B. A. Technological University, Lonere, India. Since 2002 he is working with Government College of Engineering, Directorate of Technical Education, Government of Maharashtra, India as an Assistant Professor in Instrumentation. His areas of interest include process control, computational intelligence.

Sanjay R. Patil is a member of Instrument Society of India, Biomedical Society of India and the Indian Society for Technical Education (India).

A. A. Ghatol was born on 29th August, 1948. He is presently the vice-chancellor of Dr. Babasaheb Ambedkar Technological University, Lonere-Raigad, Before joining as a vice-chancellor, he was a principal/director at College of Engineering, Pune during 2001-2005 and the principal at Government College of Engineering, Amaravati during 1994-2001. He holds a B.E. (electrical), Nagpur University, M.Tech. (electrical) and Ph.D in 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 a 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.