Diagnosis of Breast Cancer by Modular Neural Network

Harsh Vazirani
Soft Computing and Expert System Laboratory,
Indian Institute of Information Technology and Management Gwalior,
Gwalior, Madhya Pradesh-474010, India
harshiiitmg@gmail.com

Rahul Kala
Soft Computing and Expert System Laboratory,
Indian Institute of Information Technology and Management Gwalior,
Gwalior, Madhya Pradesh-474010, India
rahulkalaiiitm@yahoo.co.in

Anupam Shukla
Soft Computing and Expert System Laboratory,
Indian Institute of Information Technology and Management Gwalior,
Gwalior, Madhya Pradesh-474010, India
dranupamshukla@gmail.com

Ritu Tiwari
Soft Computing and Expert System Laboratory,
Indian Institute of Information Technology and Management Gwalior,
Gwalior, Madhya Pradesh-474010, India
rt_twr@yahoo.co.in

Citation: H. Vazirani, R. Kala, A. Shukla, R. Tiwari (2010) Diagnosis of Breast Cancer by Modular Neural Network. *Proceedings of the Third IEEE International Conference on Computer Science and Information Technology*, Chengdu, China, pp 115-119.
Final Version Available At: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5564054

***Abstract*— Diagnosis of diseases is well known problem in the medical field. Past research shows that medical database of disease can be train by using various neural network models. Many medical problems face the problem of curse of dimensionality due to the excessively large number of input attributes. Breast cancer is one such problem. We propose the use of modular neural network for effective diagnosis. In the proposed methodology four modules are made; each module gets half the problem attributes which are trained and tested by two neural network models, Back Propagation Neural Network (BPNN) and Radial Basis Function (RBFN). Integration is done using a probabilistic sum rule. The modular neural network gave an accuracy of 95.75% over training data and 98.22% over testing accuracy, which was experimentally determined to be better than monolithic neural networks.**

Keywords- Medical Expert System, Modular Neural Network, Diagnostic System, Back Propagation Neural Network, Radial Basis Function Network, Breast Cancer.

I. INTRODUCTION

There is a huge research going on into the medical field in devising automated intelligent systems for medical diagnosis. The right treatment of the patient depends on the correct diagnosis of the disease. This problem can be resolved by the Neural Networks.

The big problem with the medical dataset is the large number of attributes. The diagnosis method takes time to analyze the large number of data attributes and give the output. A problem with large dimensions results in a highly dimensional and complex feature space, and any system finds it excessively hard to construct decision boundaries, which assume complex shapes, across the various classes in such a scenario. A better efficiency may be obtained through more optimum method for diagnosis by introducing modularity at the attribute level in these systems. The approach use in this paper is to develop a decision support system which can train and test the dataset of higher dimensionality in a computational efficient manner.

The diagnosis system usually makes use of historical data to formulate rules regarding the analysis. The historical data is a collection of large amount of information regarding the presence or absence of disease in multiple scenarios. The system tries to extract valuable information or rules from this database by the process of machine learning. It then tries to generalize the extracted information or rules to the unknown data. The entire knowledge of the system to detect the disease is based on the learning of the historical data. Hence it can effectively predict the presence or absence of some disease if the behavior was recorded by the system sometimes earlier and presented at the time of learning as the historical database. In case the same does not hold, the system might not behave as desired and it may give any random or abrupt output. Hence ideally a larger historical database is preferred that can capture as much diversity as possible. The aim is that in the future any unknown input is related to one of the inputs in the historical database.

The medical dataset is used for training of the intelligent system, to find out the presence or absence of the disease. The diagnostic result is based on the input attributes. The expert system analyses and gives decision on the basis of these attributes. The efficiency of the expert system is changed with a change of these attributes. Every expert system has different efficiency with different computational time. The neural expert outputs may differ development of decision support system is of different training time and efficiency with different number of neurons involve at the hidden layer.

The Artificial Neural Networks (ANNs) form good means of learning from the past data (or machine learning) and generalizing the learnt trends into the unknown inputs for the Breast Cancer Disease. These networks hence undergo two separate stages of training and testing. One of the chief ways of training of the ANNs is Back Propagation Algorithm (BPA). The algorithm however many a times gets trapped in local minima.

The outputs from different modular neural network applications lead to the general evidence that the use of modular neural networks implies a significant learning improvement comparatively to a single neural network and especially to the Back Propagation Neural Network (BPNN).

A large amount of research in numerous problem domains is done in the past few years. In [1, 2, 3] modular neural network is used for the biometric recognition. The authors integrated the information of the three biometric recognition systems of the person i.e. face, fingerprint and voice recognition. Local experts and an integrated unit are two components which are used in the architecture. The response integration is performed using type-2 fuzzy logic which shows the improvement in the performance over type-1 fuzzy logic.

The MHC-II binding [4, 5] is predicted by the modular neural network which outperforms over the single neural network. The authors show most ANNs based prediction methods adopt a single training algorithm which is a major drawback resulting in slow learning, over-fitting and converging toward a local minimum. Modular neural network offer an advantage over monolithic ones, such as increased learning speed, improved generalization and reliability.

The ray tracing algorithm [7] was implemented by the modular neural network for radio wave propagation modeling. The modular neural network is constructed by decomposing the brute force ray tracing algorithm into small modules.

The modular neural network was applied for the speaker identification [11]. The speaker recognition, which can be classified into identification and verification, is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves.

The approach used in [12] is to improve handwritten digit recognition. In this they used two basic kind of modular networks are used. In the first kind the digits 0, 1, 2, 5, 6, 7 are the digits are provided to the seven expert modular networks. The pair of digits 3-8 and 4-9 respectively is provided for the other modular networks.

In this paper we propose to solve the problem of curse of dimensionality in a redundant 'mixture of experts' mode by using modularity at the attribute level. The complete dataset is divided into two parts, each part with half the total number of attributes in the system. Each part is trained and tested by two experts, using Back Propagation Neural Network and Radial Basis Function Network. This forms four modules of the modular neural network. The network uses probabilistic sum integration for final decision making.

II. BACKGROUND

In this section background of the paper is discussed. This section is divided into three parts Back Propagation Neural Network (BPNN), Radial Basis Function Network (RBFN), and Modular Neural Network (MNN).

A. *Backpropagation Neural Network (BPNN)*

The Back Propagation Neural Network (BPNN) consists of an input layer, one or more hidden layer, and an output layer as shown in fig. 1. The BPNN trains the neural network by propagating the error from the last layer to the first layer. All the input neurons at the hidden layer transfer their computed results to the next layer and hence the final results are propagated to the output layer. The final results at the output layer have some error according to which the weights are updated so that the error at the output is minimized.

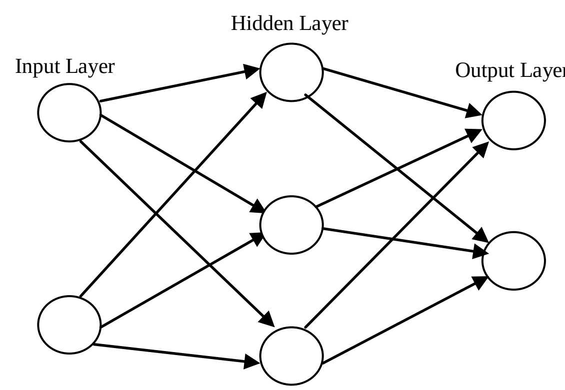

Fig.1: Network Architecture of Back Propagation Neural Network (BPNN)

This neural network is used in many domains one of them is recognition problems such as voice, speech, face etc. The other domain is the medical dataset where it is applied to medical diagnosis.

Backward error propagation, or simply Back Propagation, is the most popular learning algorithm for connectionist learning. As the name implies, an error in the output node is corrected by back propagating this error by adjusting weights through the hidden layer and back to the input layer. While relatively simple, convergence can take some time depending upon the allowable error in the output.

B. Radial Basis Function Network (RBFN)

In this the whole network consists of an input layer, one hidden layer, and an output layer. This neural network has nonlinear transfer function at the hidden layer. Each of these three layers has different tasks. All the nodes are connected to each other the final output is linear. The Gaussian transfer function is used by the neurons at the hidden layer whose output is inversely proportional to the distance from the center of the neuron.

Radial Basis Function uses the radial functions as activation function. This forms the linear combination of the radial basis functions. The advantage of using RBFN is the training speed, taking into account that this process involves, usually two distinct stages: an unsupervised training and a supervised training. In the unsupervised training the centers are created at intermediary layer. Commonly, this stage employs k-means algorithm. In supervised training, a linear method is employed to minimize the established error measure. Changing parameters of one neuron has only a small effect for input values that are far away from the center of that neuron.

C. Modular Neural Network (MNN)

There exists a lot of neural network architectures in the literature that work well when the number of inputs is relatively small, but when the complexity of the problem grows or the number of inputs increases, their performance decreases very quickly. For this reason, there has also been research work in solving some way the problems in learning of a single neural network over high dimensional spaces [3].

The capabilities of ANN in some cases when used are not satisfactory like in the case of large dataset. The modular neural network is used in such cases which work as a combination of neural networks. The modular neural network has a hierarchical organization comprising multiple neural networks. The combination of estimators may be able to exceed the limitation of a single estimator. The idea also shares conceptual links with “divide and conquer” methodology [8, 9, 10].

Modular artificial neural networks are especially efficient for certain classes of regression and classification problems, as compared to the conventional monolithic artificial neural networks. These classes of problems include problems that have distinctly different characteristics in different operating regimes. For example, in the case of function approximation, piece wise continuous functions cannot in general be accurately modeled by monolithic artificial neural networks.

But on the other hand, modular neural networks have proven to be very effective and accurate when used for approximating these types of functions. Some of the main advantages of learning modular systems are extendibility, incremental learning, and continual adaptation, economy of learning and re-learning, and computational efficiency.

III. METHODOLOGY

The proposed modular neural network is applied over the problem of Breast Cancer diagnosis. The data for the problem is taken from UCI Machine Learning Repository [13]. This dataset contains the 30 attributes and 569 instances.

The first problem that the paper addresses is curse of dimensionality. An individual neural network may not be able to effective solve the problem due to a large number of attributes (30) in it. Hence we divide the complete database into two parts of equal number of attributes. In this manner we insert modularity in attributes for the diagnosis. The two sub-problems, having limited number of attributes may be effective in achieving a decent diagnostic score with less complex feature space. The anomalies in one sub-problem may be removed by the other sub-problem. In all we divide one high dimensional feature space, into two spaces with half the dimensionality. Some classes’ difficult to separate in one sub-space may be easily done in the other feature space. This removes problem of curse of dimensionality without a big loss in information available to the system for diagnosis.

The other problem that the paper addresses is of sub-optimal performance by a single neural network model. The limitations of the individual models may prohibit effective construction of the decision boundaries. We hence make use of multiple neural network models to solve the same problem in a redundant manner. These neural networks act as experts to construct the decision boundary to the best of their abilities. The decision boundaries to the complete system may be formulated using these individual decision boundaries. In this paper we use BPNN and RBFN as the two experts.

Using both attribute division, and mixture of experts we get a total of 4 modules. The attributes from 1-15 is given to the module 1 and module 2. Attributes from 16 to 30 is given to the module 3 and module 4. The module1 and module 3 are trained and tested with the BPNN. The module 2 and module 4 are trained and tested with the RBFN.

Each module gives as its output the probability vector denoting the probability of occurrence of the disease. This is a value between 0 and 1, 0 denoting absence of disease and 1 denoting the presence of disease. Each module is associated with a weight denoting the contribution of the module in the decision making of the complete system. This is a real number externally set by the designer keeping in mind the system performance.

The various modules give the outputs to the integrator. The integrator does the task of making the final diagnostic decision, considering all the individual responses of the experts. Let the results obtained from module 1, module 2, module 3, and module 4 be O_1, O_2, O_3 and O_4. Here any output O_i denotes the probability of occurrence of breast cancer as observed by module i. Further let w_1, w_2, w_3, and w_4 be the weights associated with these 4 experts. The output obtained after integration from the integration unit is given by (1).

$$O = o_1 w_1 + o_2 w_2 + o_3 w_3 + o_4 w_4 \qquad (1)$$

Where $w_1 + w_2 + w_3 + w_4 = 1$,

If the integrated output O is greater than 0.5 then the breast cancer is classified as malignant in the patient and if O is less than 0.5 then the cancer is classified as benign. This is given by (2).

$$c = \begin{cases} benign & o \geq 0.5 \\ malignant & o < 0.5 \end{cases} \tag{2}$$

Here c is the final diagnostic result of the system.

The complete architecture of the system is shown in fig. 2. The figure depicts the division of attributes and further the use of different neural network for each attribute division. The integrator does the final task of integrating the individual neural responses by the various modules. Fig. 3 shows the integrator which gives the final output.

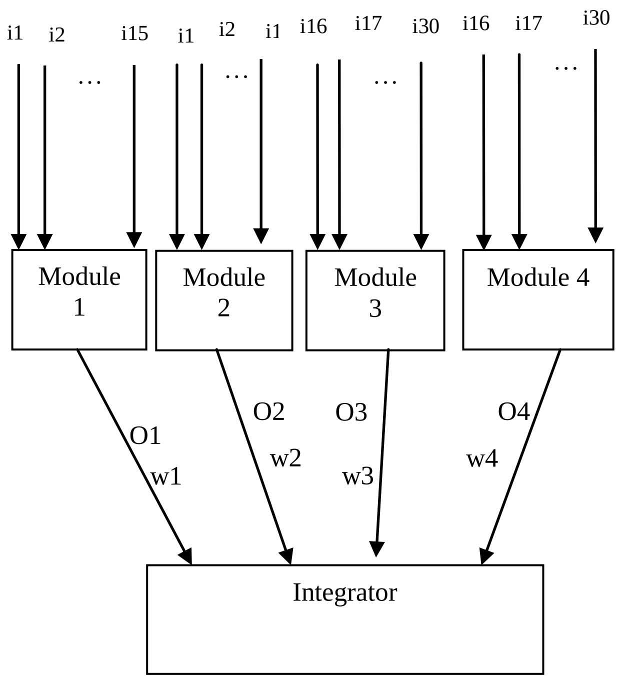

Fig. 2: General Architecture of the Modular Neural Network

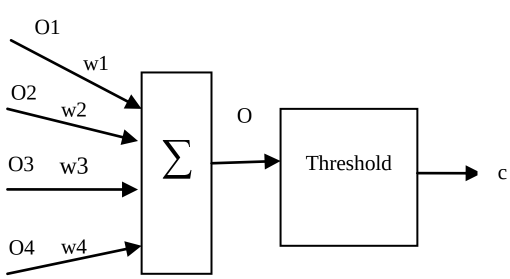

Fig. 3: Architecture of Integrator

The general methodology used in this paper starts from the collection of dataset of breast cancer from the UCI machine learning repository [13]. The dataset is divided into 70% training data and 30% training data. The training and testing is performed by the BPNN and RBFN in different modules after the division of attributes among the modules. The general methodology is shown in the fig. 4.

IV. RESULTS

The two neural network models are used for the experiment. The first method used for the problem was BPNN. MATLAB was used as a platform for the implementation. Here we used a single hidden layer which consisted of 25 neurons. The activation functions for the hidden layer was *tansig* and *purelin*. The training function used was *traingd*. The other parameters were a learning rate of 0.05 and a goal of 10^{-1}. Training was done till 2000 epochs.

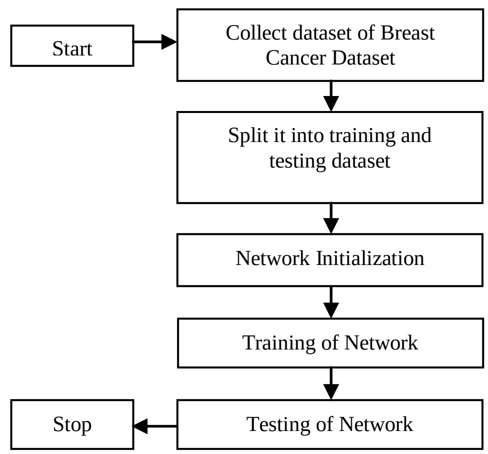

Fig.4: Overview of working of the methodology

Ideally the network must always output 0 or 1 denoting the absence or presence of the disease. However this may not be practically possible due to numerous reasons. The system always gives continuous outputs. The output is closer to 1 for benign cases and closer to 0 for the malignant cases. We place a threshold here that helps in the decision of the presence or absence of diseases. Any value of the network above this threshold is taken to be 1 and values less than this threshold are taken to be 0. In this problem we fix the threshold as 0.5.

After the network was trained and tested with BPNN, the performance of the system for module1 was found to be 89.50% for the training data set and 96.44% for the testing data set with elapsed time of 3.88 seconds. The performance of the system for module 3 was found to be 91.5% for training data and 94.67% for testing data with elapsed time of 3.82 seconds. The performance of the system with all the 30 attributes was found to be 91% for training data and 96.44% for testing data and elapsed time is 5.58 seconds.

The second method was RBFN. MATLAB was used as a platform for the implementation. Again the threshold is set to

0.5. If the value comes out to be is greater than this threshold the output is malignant else the output is benign.

The performance after the network was trained for module2 are 94.75% for the training data and 96.44% for testing data. For the module 4 the performance was 97.50 % for training data and 97.63% for the testing data. The performance for the whole dataset of 30 attributes was found to be 97.75% for training data and 97.63 % for the testing data.

The summary of the experiments is given in Table 1. It may be clearly seen that the modular neural network gave accuracy greater than both the monolithic neural networks i.e. BPNN and RBFN. The testing accuracy of RBFN was recorded to be 97.63%. This was better than that of the BPNN which gave am accuracy of 96.44%. However the proposed architecture gave the best performance of 98.22%. This proves that the dimensionality was a problem in the database, which was solved using the proposed modular neural network approach.

V. CONCLUSION

In this paper we proposed method of diagnosis of breast cancer disease by the use of Modular Neural Networks. Two different neural network models were used in a mixture of experts' architecture. These two networks models performed independently on four different modules in solving the problem. The results from these are integrated by the use of integrator. The integrator adopted a probability based approach in trying to figure out the final solution to the problem.

Table 1: Experimental Results

Module	Method	Attributes	Training Accuracy	Testing Accuracy	Time (secs)
1	BPA	1-15	89.50%	96.44%	3.88
2	RBFN	1-15	94.75%	96.44%	0.25
3	BPA	16-30	91.50%	94.67%	3.82
4	RBFN	16-10	97.50%	97.63%	0.29
-	**MNN**	**1-30**	**95.75%**	**98.22%**	**8.24**
-	BPNN	1-30	91 %	96.44 %	5.58
-	RBFN	1-30	97.25%	97.63%	0.25

In this paper we proposed the use of modular neural network for breast cancer diagnosis. The entire system first divided the attributes into sets of smaller attributes to escape from the curse of dimensionality, which is a known problem in neural computation. Each problem set was solved using two different neural models. Probabilistic sum integration was used to get the final result. The approach was applied over the problem of diagnosis of breast cancer. It was observed that the system so evolved was easily able to solve the problem and gave an accuracy that was much better than the monolithic neural methods. Hence this can be an effective approach for disease diagnosis.

The approach discussed was applied over the problem of breast cancer disease diagnosis. This database is not a very large in terms of the number of attributes and the number of data instances. The experimentation over the other databases with very large number of attributes and number of data instances may be done that may give more efficient results with the use of modular neural networks.

ACKNOWLEDGMENT

This work is sponsored and supported by Indian Institute of Information Technology and Management Gwalior.

REFERENCES

[1] Jerica Urias, Denisse Hidalgo, Patricia Melin, and Oscar Castillo, "A New Method for Respose Integration in Modular Neural Networks using Type-2 Fuzzy Logic for Biometric Systems", Proceedings of International Joint Conference on Neural Networks, Florida, USA, August 12-17, 2007

[2] Jerica Urias, Patricia Melin, and Oscar Castillo, "A Method for Respose Integration in Modular Neural Networks using Interval Type-2 Fuzzy Logic for Biometric Systems", Analysis and Design of Interlligent Systems using Soft Computing Techniques, Volume 41/2007, PP 5-15, Springer Berlin / Heidelberg , 2007

[3] Jerica Urias, Patricia Melin, and Oscar Castillo, "Optimization with Genetic Algorithms of Modular Neural Networks using Interval Type-2 Fuzzy Logic for Response Integration: the case of Multimodal Biometry", Bio-inspired Hybrid Intelligent Systems for Image Analysis and Pattern Recognition, Volume 256/2009, PP 111-121, Springer Berlin / Heidelberg , 2009

[4] An Zeng, Qi-Lun Zheng, Dan Pan, and Hong Peng, "Utilizing Modular Neural Network to Predict MHC Class II-binding Peptides", Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2004

[5] Gasser Auda, Mohamed Kamel, and Hazem Raafat, "Modular Neural Network Architectues for Classification", Proceedings of IEEE International Conference on Neural Networks, ICNN, 1996

[6] L. K. Hansen, Y. P. Salomon, "Neural Network Ensembles", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990.

[7] J.W.H. Lee, A.K.Y. Lai, Qi-Lun Zheng, "Function-Based and Physics-Based Hybrid Modular Neural Networks for Radio Wave Propagation Modelling", Proceedings of IEEE Antennas Propag Soc AP S INT Symp, 2000

[8] Vineet R. Khare, Xin Yao, Bernhard Sendhoff, Yaochu, Heiko Wersing, "Co-evolutionary Modular Neural Networks for Automaric Problem Decomposition", Congress on Evolutionary Computation (CEC). IEEE Press, pages 2691-2698, 2005

[9] J.W.H. Lee, A.K.Y. Lai, Qi-Lun Zheng, "Function-Based and Physics-Based Hybrid Modular Neural Networks for Radio Wave Propagation Modelling", Proceedings of IEEE Antennas Propag Soc AP S INT Symp, 2000

[10] Yu, B., He X., "Training Radial Basis Function Networks with differential evolution", Proceedings of IEEE International Conference on Granular Computing, Atlanta,GA, USA, 2006

[11] Gabriela Martinez, Patricia Melin, and Oscar Castillo, "Optimization with Genetic Algorithms of Modular Neural Networks using Hierarichcal Genetic Algorithm Apllied to Speech Recognition,

Prcoceedings of International Jont Conference on Neural Networks", Canada, 2005

[12] Claudio A Perez, Patricio A Galdmas, and Carlos A Holzmann, "Improvements On Handwritten Digit Recognition By Cooperation Of Modular Neural Networks", IEEE International Conference On Systems Man And Cybernetics, VOL 5, pp. 4172-4177, USA, 1998. *)*

[13] William H. Wolberg, W. Nick Street, Olvi L. Mangasarian, UCI Machine Leaning Repository, [http://www.ics.uci.edu/~mlearn/MLRepository.html], University of Wisconsin, 1995, Available at : http://archive.ics.uci.edu/datasets/ Breast+Cancer+Wisconsin+(Diagnostic)