Available at: www.researchpublications.org

Neural Network based Decision Support System for the Diagnosis of Thyroid Diseases

Ms.Wrushali Mendre[1], Dr.Ranjana D.Raut[2]

[1] *Department of Electronics and Telecommunication Engineering, SKNCOE, Pune University, Pune India*
[2] *Departments of Applied Electronics, SGB Amravati University, Amravati, India*

[1]vrushalimendre@gmail.com
[3]rdr24164@rediffmail.com
[1]*Smt.Kashibai Navale College of Engineering*
STES Vadgaon Campus,
S.No.44/1
Pune-411041

***Abstract*— Aim of this research work is to investigate the ability of neural network to discriminate the two subtypes, hypothyroid and negative form, of thyroid disease, on the basis of laboratory clinical data base. The optimal parameters are identified for the suggested neural networks (NN) like Multi Layer Perceptron (MLP), Radial Basis Function (RBF) and Principal Component Analysis (PCA). The three optimal NN models are chosen, which are tested exhaustively on the available data sets for classification of two forms of thyroid disease such as hypothyroid or negative. Different methods are used for data partitioning, and rigorous experimentation is carried out on different data sets in order to prove the efficacy and robustness of NN.**

***Keywords*— Multi layer Perceptron, Neural Network, Radial Basis Function, Principal Component Analysis, Classification Tree**

I. INTRODUCTION

Thyroid hormone produced by the thyroid gland helps regulate the body's metabolism. The thyroid data are obtained from the measurements of the thyroid gland. The thyroid makes two active thyroid hormones, levothyroxine (abbreviated T4) and triiodothyronine (abbreviated T3). These hormones are important in the manufacture of proteins, in the regulation of body temperature, and in overall energy production and regulation [1].

In general, thyroid disease can be divided into two broad groups of disorders: those which primarily affect the function of the thyroid and those which involve neoplasms, or tumors of the thyroid. Both types of disorders are relatively common in the general population. Most thyroid problems can be treated successfully. Abnormalities of thyroid function are usually related to production of too little thyroid hormone (hypo-thyroidism) or production of too much thyroid hormone (hyperthyroidism). Hypothyroidism, or an under active thyroid, has many causes. Some of the causes are prior thyroid surgery, exposure to ionizing radiation, chronic inflammation of the thyroid (autoimmune thyroiditis), iodine deficiency, lack of enzymes to make thyroid hormone, and various kinds of medication [2].

Hyperthyroidism, or an overactive thyroid, may also be caused by inflammation of the thyroid, various kinds of medications, and lack of control of thyroid hormone production. One of the most common causes is Graves' disease. Graves' disease happens when the body makes proteins that constantly tell the thyroid to make more thyroid hormone [3].

Proper interpretation of the thyroid data besides clinical examination and complementary investigation is an important issue on the diagnosis of thyroid disease. In recent times, neural networks have been employed as a widely used method for designing Decision Support System (DSS) for disease diagnosis. There has been a considerable research going on in developing computer based decision support systems for improving the quality of health care. Two problems are the most common in the field of automatic diagnosis: the selection of necessary parameter set for right diagnosis and forming of steady and powerful algorithm which doesn't require long time to run.

There have been efforts taken in last few years to develop artificial neural network based models for clinical applications. Anupama, et al [4] have proposed ANN based models successfully. ANN based models have advantage over other types of decision support systems, as ANN based systems are self organizing type and have learning capabilities.

The paper is organized as follows. First the optimal MLP NN based DSS is designed to diagnose the given Thyroid data base. Later, the RBF and PCA neural network based DSS are developed for the binary classification task. The exhaustive data partitioning is done with a view to test the proposed NN for their robustness as classifier.

II. ANN CLASSIFIERS

The performances of three different networks in classification of thyroid data sets are compared in this article. Those include Multilayer Perceptron (MLP), Radial-Basis Functions (RBF) and Principal Component Analysis (PCA). A brief description of them follows.

A. MLP networks:

MLPs are feed-forward neural networks trained with the standard back-propagation algorithm. It is shown that a network having a single layer of threshold units could classify a set of points perfectly if they were linearly separable [5]. It is shown that for a set of N data points, a two-layer network of threshold units with $N-1$ units in the hidden layer could exactly separate an arbitrary dichotomy [6]. Since it is very likely that one ends up in a "bad" local minimum, the network should be trained a couple of times (typically at least five times), starting from different initial weights.

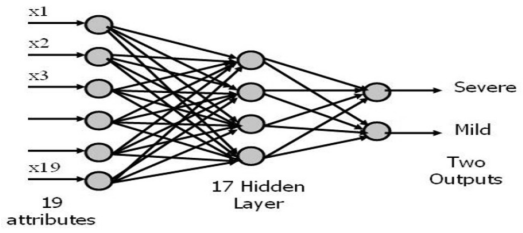

Fig.1. Multi Layer Perceptron NN

As it has been pointed out in earlier discussion, a MLP NN shown above in fig.1 is chosen as a DSS. In order to design a proper architecture of the MLP NN model; a computer simulation experiment is designed where the number of hidden neurons is varied gradually from 1 to 20. It is found that the performance of the selected model is optimal for 17 neurons in the hidden layer with regard to the MSE, NMSE, correlation coefficients, area under the ROC curve, and percent classification accuracy for the testing data set. When we attempted to increase the number above 17 in the hidden layer, the performance of the DSS was not seen to improve.

TABLE I: Optimal Parameters for MLP NN DSS

S.N.	Parameter	Hidden Layer#1	Output Layer
1	Processing Elements	17	2
2	Transfer Function	Lin Tanh	Soft-max
3	Learning rule	Momentum	Delta Bar Delta

The total number of free parameters (connection weights) for NN model (19-17-2) is 376 (19 X 17 + 17 X 2 + 17 +2). Now this NN model is trained (three times with different random weight initialization) with 5000 iterations of the standard back propagation algorithm with momentum term. Table 1 indicates optimal parameters used for MLP NN based DSS. Variation of average of minimum MSEs with respect to PEs in hidden layer is graphed in Fig. 2.

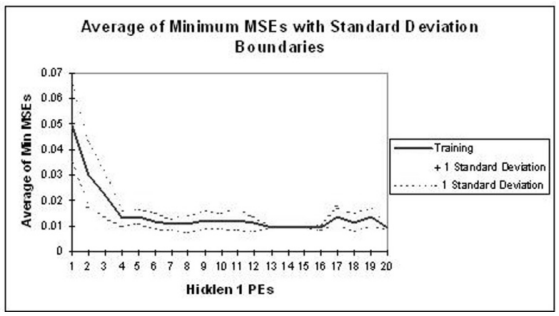

Fig. 2. Average of Minimum MSEs versus no. of PEs in the hidden layer

Max. Epochs = 5000, No. of Input PEs=19

B. RBF networks:

Radial-basis functions (RBF) were first introduced in the solution of the real multivariate interpolation problem [15]. The construction of a RBF network as shown in fig. 3, in its most basic form, involves three layers with entirely different roles. The input layer is made up of source nodes (sensory units) that connect the network to its environment.

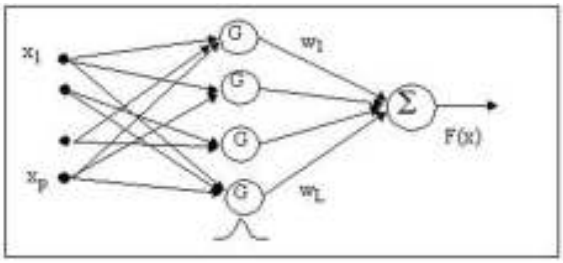

Fig.3. Radial Basis Function

Different initial conditions are tried to make sure that we are really converging to the absolute minimum. Moreover, this also removes biasing in the learning mechanism. Therefore, in order to ensure true learning and proper generalization, the network is run at least three times to gauge performance. The training of RBF constitutes 100 epochs in unsupervised learning mode for setting the centres' and width of the Gaussians and at least 1000 epochs in the supervised learning mode to compute the connection weights in the output layer. The supervised learning may terminate earlier if the minimum specified error threshold of 0.01 is reached earlier. An exhaustive experimental study has been carried out

to design the optimal parameters of RBF NN as depicted in table 2.

TABLE I II: Optimal Parameters for RBF NN

S.N.	Optimal Parameters of RBF NN	
1	No. of Clusters	120
2	Transfer Function	Softmax
3	Supervised learning rule	DeltaBarDelta
4	Competitive learning metric	Euclidean
5	Unsupervised learning rule	Conscience-full

Experimental results show that, the highest classification accuracy obtained at clusters 120.

PCA networks

Principal component analysis networks (PCAs) combine unsupervised and supervised learning in the same topology. PCA is a very well known statistical procedure in which input data of very large dimensionality (D dimensions) is projected onto a smaller-dimensionality space M (M < D), a step that is commonly called feature extraction as shown in fig.4. The linear projection that accomplishes this goal is the PCA. It is an unsupervised linear procedure that finds a set of uncorrelated features, principal components, from the input. A MLP is used to perform the nonlinear classification from these components.

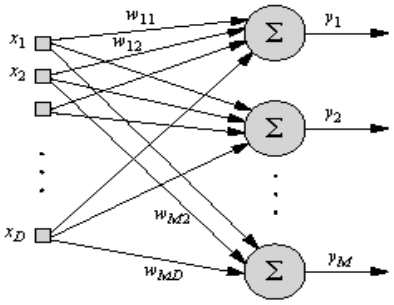

Fig. 4. Principal Component Analysis network

PCA is a data reduction method, which condenses the input data down to a few principal components. As with any data reduction method, there is the possibility of losing important input information. The two most robust learning rules are Oja's and Sanger's implementations of the Hebbian principle. Between the two, Sanger's is preferred for PCA because it naturally orders the PCA components by magnitude.

Table 2 shows, the optimal parameters. Where, D = 19 attributes and M=10 principal components (M<D). These 10 attributes are used as input for MLP reducing the free weights and complexity.

TABLE III
Optimal Parameters for PCA NN DSS

S.N.	Parameter	Hidden Layer#1	Output Layer
1	Processing Elements	17	2
2	Transfer Function	Lin tanh	Soft-max
3	Learning rule	Momentum	Delta Bar Delta

Max. Epochs = 5000, No. of Input PEs=17, Principal Components =10, Learning rule = Sanger Full.

III. NN PERFORMANCE INDEX

A. MSE (Mean Square Error)

The formula for the mean squared error is defined by following formula:

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}\left(d_{ij} - y_{ij}\right)^2}{N \cdot P}$$

B. NMSE (Normalized Mean Square Error)

The normalized mean squared error is defined by following formula:

$$NMSE = \frac{P \cdot N \cdot MSE}{\sum_{j=0}^{P}\frac{N\sum_{i=0}^{N} d_{ij}^{\,2} - \left(\sum_{i=0}^{N} d_{ij}\right)^2}{N}}$$

Where, P = number of output processing elements, N = number of exemplars in the data set, MSE = mean square error, d_{ij} = desired output for exemplar i at processing element j.

C. R (correlation matrix)

The correlation coefficient (r) solves this problem. By definition, the correlation coefficient between a network output x and a desired output d is:

$$\text{Where, } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad \bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$$

The correlation coefficient is confined to the range [-1, 1]. When r = 1 there is a perfect positive linear correlation between x and d, that is, they co vary, which means that they vary by the same amount. When r = -1, there is a perfectly linear negative correlation between x and d, that is, they vary in opposite ways (when x increases, d decreases by the same amount). When r = 0 there is no correlation between x and d,

i.e. the variables are called uncorrelated. Intermediate values describe partial correlations.

IV. RESULTS

In this paper, the performance of the three Neural Networks MLP, RBF and PCA are evaluated for classification of thyroid disease as hypothyroid or negative. The title of the data-set is thyroid gland data, taken from UCI repository of machine learning data [8]. Two types of data partitioning is used, dataset 1 with 20% samples for training and 80% samples for testing, dataset 2 with 60% samples for training and 40% samples for testing. The classification accuracy and performance of Neural Network MLP, RBF and PCA is shown in the following tables 3-9.

TABLE IV
Performance Evaluation of MLP Data Set-I
(Train 80% Test 20%)

Sr. No	Performance Index	Negative	Hypothyroid
1	Classification Accuracy	100	100
2	MSE	0.009	0.0589
3	NMSE	0.038	0.4058
4	R	0.985	0.8351

TABLE IV
Performance Evaluation of MLP Data Set-II
(Train 60% Test 40%)

Sr. No	Performance Index	Negative	Hypothyroid
1	Classification Accuracy	100	90
2	MSE	0.008	0.0589
3	NMSE	0.048	0.3058
4	R	0.885	0.7351

TABLE V
Performance Evaluation of RBF Data Set-I
(Train 80% Test20%)

Sr. No	Performance Index	Negative	Hypothyroid
1	Classification Accuracy	94.44	75.00
2	MSE	0.101	0.057
3	NMSE	0.406	0.318
4	R	0.779	0.874

TABLE VI
Performance Evaluation of RBF Data Set-II
(Train 60% Test 40%)

Sr. No	Performance Index	Negative	Hypothy.
1	Classification Accuracy	100	90.00
2	MSE	0.071	0.035
3	NMSE	0.285	0.197
4	R	0.857	0.933

TABLE VII
Performance Evaluation of PCA Data Set-I
(Train 80% Test 20%)

Sr. No	Performance Index	Negative	Hypothy.
1	Classification Accuracy	80.00	90.00
2	MSE	0.101	0.057
3	NMSE	0.406	0.318
4	R	0.779	0.874

V. CONCLUSION

From tables 1 to 9 it is observed that, percentage average Classification accuracy of MLP NN is higher as compared to RBF and PCA NNs and it has performed excellent on such complex and insufficient data set.

TABLE VIII
Performance Evaluation of PCA Data Set-II
(Train 60% Test 40%)

Sr. No	Performance Index	Negative	Hypothyroid
1	Classification Accuracy	100	100
2	MSE	0.009	0.0589
3	NMSE	0.038	0.4058
4	R	0.985	0.8351

Table IX
Comparison of MLP, RBF and PCA

Neural Network	% Average Classification Accuracy	
	Negative	Hypothyroid
MLP	**100**	**95**
RBF	97.22	87.5
PCA	95	88

REFERENCES

[1] Eystratios G. Keramidas, Dimitris K. Iakovidis, "Thyroid Texture Representation via Noise Resistant Image Features " ,IEEE International Symposium on computer – based Medical Systems , 2008, pp 560-565

[2] Chuan-Yu Chang, Ming-Fang Tsai and Shao-Jer Chen," Classification of the Thyroid Nodules Using Support Vector Machines" International joint conference on Neural Networks 2008, pp 3093- 3098.

[3] Jianrui Ding, H.D. Cheng, Chunping Ning, "Novel Quantitative Measurement for Thyroid Cancer Detection Based on Elastography", The 4th International congress on Image and signal processing. 2011, pp-1801-1804.

[4] D.E. Maroulis, M.A. Savelonas and S.A. Karkanis, "Computer –Aided Thyroid Nodule Detection in Ultrasound Images" 18th IEEE International Symposium on computer based medical systems(CBMS'05),2005.

[5] C.C.Leung, F.H.Y. Chan, K.Y. Lam and P.C.K. Kwok, "Thyroid Cancer Cells Boundary Location by Fuzzy Edge Detection Method" , 2000 IEEE. pp 360 – 363.

[6] M. Savelonas, D. Maroulis, D. Iakovidis and S. Karkanis, "Variable Background Active Contour Model for Automatic Detection of Thyroid Nodules in Ultrasound Images ", 2005.

[7] Gheorghe Gavriloaia and Roxana-Mariuca Gavriloaia, "Spatiotemporal Filtering for Ultrasound Images", International joint conference on biomedical engg. And informatics,2011, pp 502- 505.

[8] UCI repository of machine learning databases,
http.//archive. ics. uci. edu/ml/datasets. html accessed November