IMPLEMENTATION OF NEURO FUZZY CLASSIFIER MODEL FOR THYROID DISEASE DIAGNOSIS

Ms. Siddhi Vinod Parodkar[1], Ms. Amita Dessai[2]

[1,2]*Department of Electronics and Telecommunication, Goa College of Engineering (India)*

ABSTRACT

Thyroid diseases are rising at an alarming rate worldwide across several age groups. Current practice for thyroid disease diagnosis is doctor's examination and a number of blood tests which requires lots of experience and knowledge. Because of the difficulty in considering large number of interrelated measurements and non-specific nature of symptoms of a disease, experts often have imprecise knowledge which causes uncertainty in diagnosis procedure. Thyroid disease detection via proper interpretation of the thyroid data is an important classification problem. The main objective of this project is to design an intelligent system that can diagnose the thyroid diseases with minimum diagnosis time and enhanced diagnosis accuracy.

Keywords: Feature Selection (FS), Linguistic Hedge (LH), Neuro-Fuzzy Classifier (NFC), Scaled Conjugate Gradient algorithm (SCG), Thyroid Disorders.

I. INTRODUCTION

The thyroid is an endocrine gland which absorbs iodine from food, which is converted into thyroid hormones: thyroxin (T4) and triiodothyronine (T3). The produced hormones are released into the bloodstream to all the other organs which are crucial for body's metabolism and growth. Thyroid diseases are emerging as one of the most common endocrine disorders worldwide. Thyroid diseases differ from other diseases in terms of their ease of diagnosis, accessibility of medical treatment, and the relative visibility. Both the rise and decline in thyroid hormone secretions can cause health problems. Abnormalities of thyroid hormones are divided into two categories: Hypothyroidism and Hyperthyroidism. Hypothyroidism is caused when thyroid hormones are produced in lesser quantities than in its normal condition whereas Hyperthyroidism is caused when production of thyroid hormones is in excess than in its normal condition. Normal functioning condition of thyroid gland is called Euthyroid. Accurate diagnosis of the thyroid gland disease based on tests, signs and symptoms is a very challenging task. A series of clinical observations are required to be undertaken to detect the presence of a particular disease, often causing the delay of a correct diagnosis decision. Thus, the diagnosis of thyroid gland disease is complicated due to many factors. For instance, the clinical tests, signs and symptoms of thyroid gland disease are associated with many human organs other than the thyroid gland, and very often thyroid diseases may exhibit various syndromes. At the same time, different types of thyroid diseases may have identical symptoms. Thus, thyroid disease diagnosis via proper interpretation of the thyroid data is an important classification problem in medical field. Hence, disease diagnosis at early stages with higher accuracy is desired. Significant lifesaving of patients suffering from thyroid diseases can be achieved, if an accurate diagnosis

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

decision can promptly be made; after which doctors can immediately plan an appropriate future treatment. The patterns that are hidden among vast amounts of data in the hospital databases; specifying test results, signs and symptoms; can be used for classifying diseases.

The main objective of this research work is to design a neuro-fuzzy rule based classifier to classify thyroid disorders into three types- Hypothyroid, Hyperthyroid and Euthyroid by extracting more significant features (prescribed measurements) from database.

II. PROPOSED METHODOLOGY

This section gives brief description about the proposed algorithm for classifying thyroid diseases into three classes. Fig.1. shown below is the flowchart for neuro-fuzzy approach.

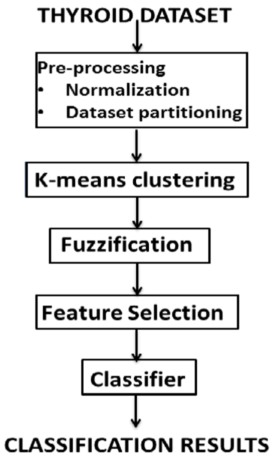

Fig.1. Flowchart of proposed neuro-fuzzy approach.

The following algorithm is developed to implement proposed Neuro-Fuzzy Classifier to determine whether a patient has a normally functioning thyroid (euthyroid), an under-functioning thyroid (hypothyroid), or an overactive thyroid (hyperthyroid).

Algorithm:

Step 1: Thyroid dataset collection.

Step 2: Pre-prossessing the dataset by normalizing input features using standardization or Z-score method.

Step 3: Partitioning the dataset into training set and testing set using Hold-out partitioning method.

Step 4: K-means clustering to partition dataset according to three classes.

Step 5: Fuzzification of input features.

Step 6: Feature Selection using Linguistic Hedge method.

Step 7: Designing Neuro-fuzzy classifier

(a) Training classifier using scaled conjugate gradient (SCG) learning algorithm

(b) Testing the classifier

Step 8: Evaluation of performance parameters.

Step 9: Developing user-friendly GUI.

1.1. Dataset Description

The thyroid dataset used for this research work is collected from UCI Repository [13]. Dataset consists of diagnostic data records of 7200 patients related to thyroid dysfunction. For each of the 7200 cases, there are 21 attributes (6 continuous and 15 binary variables) and a class attribute (1=euthyroid (normal), 2=hyper-thyroid, 3=hypo-thyroid) used to determine in which of the three classes the patient belongs. Table 1 lists these 21 input variables and their descriptions. These attributes represent information about the results of various prescribed clinical tests carried out on a patient, signs and symptoms of disease, other health condition and general queries related to thyroid function. These attribute values have been taken as input to classify thyroid gland diseases in three different classes. The euthyroid class represents 2.3% (166 cases) of the data samples, the hyperthyroid class accounts for 5.1% (368 cases) of the observations, while the hypothyroid group makes up the remaining 92.6% (6666 cases). This highly unbalanced data set is a notoriously difficult problem for traditional classification methods.

Table1. Description of variables in thyroid dataset.

SN	Feature Name	Description
1	Age	Patient age in years.
2	TSH	TSH (Thyroid-Stimulating Hormone) test results.
3	T3	T3 (Triiodothyronine) test results.
4	TT4	TT4 (Thyroxin) test results.
5	T4U	T4U test results.
6	FTI	FTI calculated from TT4 and T4U values.
7	Sex	Patient’s gender.
8	On Thyroxine	Patient on thyroxine treatment.
9	Query on Thyroxine	Patient thyroxine treatment status unknown or unreported.
10	On Antithyroid	Patient is on antithyroid medication.
11	Sick	Patient reports illness.
12	Pregnant	Weather patient is pregnant.
13	Thyroid surgery	Patient has history of thyroid surgery.
14	I131 treatment	Patient is currently receiving iodine 131 treatment.
15	Query Hypothyroid	Patient responses indicate likelihood of hypothyroidism.
16	Query Hyperthyroid	Patient responses indicate likelihood of hyperthyroidism.
17	Lithium	Patient is on lithium treatment.
18	Goitre	Patient has goiter.
19	Tumor	Patient has thyroid tumor.
20	Hypopituitary	Patient is hypopituitary.
21	Psych	Patient has psychological Symptoms.
Class Attribute	Class	1=Euthyroid(Normal), 2=Hyperthyroid, 3=Hypothyroid.

1.2. Normalization of Dataset

Standardization or Z-score method is used for normalization of dataset. This method preserves range (maximum and minimum) and introduces the dispersion of the series (standard deviation / variance). If data follow a Gaussian distribution, they are converted into a Normal distribution and the comparison between series (probabilities calculation) will be easier. It converts all data points of particular feature to a common scale with an average of zero and standard deviation of one.

1.3. Dataset Partitioning

Hold-out Partitioning is used for partitioning thyroid dataset into training and test sets in desired ratio. In this research work thyroid dataset is partitioned in (0.7:0.3) ratio. That is, training set consists of 70% (5040 samples) of dataset and test set comprises of remaining 30% (2160 samples) of dataset.

Hold-out method has advantage over other method since it yields independent train and test sets; which helps to find robustness of the model to unknown samples. It also partitions highly un-proportionate datasets such that training and test sets will have equal proportion of samples of each class. For example, Training set consists of 70% class 1 samples, 70% class 2 samples, 70% class 3 samples, and testing set contains 30% class 1 samples, 30% class 2 samples, 30% class 3 samples.

1.4. K-means Clustering

K-means clustering aims to partition n observations belonging to each input feature into K clusters; in which each observation belongs to the cluster with the nearest mean; where K represents total number of classes. It maps input into fuzzy sets or regions. In this project, each input feature is mapped into k=3 clusters. Once clusters are formed using K-means clustering, centroid of the clusters is computed which is used as initial center of Gaussian MF before training.

1.5. Fuzzification of input features

Fuzzification is the process of making a crisp quantity fuzzy. A membership function (MF) is a curve that defines how each data point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. Membership functions (MF) are generated by training algorithm of ANN which explores the information of different features for each pattern and collects the hidden or interrelated information. Thus, in fuzzification, these membership functions are used to fuzzify crisp input feature vector; to provide better classification accuracy. Fuzzification determines the degree of membership to which these crisp inputs belong to each of the appropriate fuzzy sets. There are numerous methods to assign membership values or functions to fuzzy variables. This assignment process can be instinctive or it can be based on some algorithmic or logical operations. The method used in this research work to create fuzzy membership functions for fuzzy classes of an input data set is by using training algorithm of ANN. In this method the feature wise belongingness of the pattern to each given class is determined. For this the backpropagation neural network is trained for each feature vector as the input and expected class as the output. After training, it gives the membership value to each attribute value with respect to each class.

1.6. Feature Selection Using Linguistic Hedge Method

Feature selection is used for dimensional reduction of dataset and to obtain better feature space. Feature selection helps in extraction of patterns from complicated or imprecise data by eliminating irrelevant features for classification. The values of LHs can be used to show the importance of fuzzy sets. When this property is used for classification problems, and when a fuzzy classification rule is used to define every class, the LHs of every fuzzy set denote the importance of input features. If the LHs values of features are close to concentration values, these features are more crucial or relevant, and can be selected. On the contrary, if the LH values of features are close to dilation values, these features are not crucial, and can be eliminated. Thus, the redundant, noisy features can be eliminated, and significant features can be selected according to the LH value of features. Hence, the powers of fuzzy sets are used for feature selection. In this technique, if linguistic hedge values of class for particular feature are greater than 0.5 and close to 1, this feature is relevant, and otherwise it is irrelevant.

A feature selection and rejection criterion is created by using power values of features. There are two selection criteria, first criteria is the selection of features that have the highest hedge value 'for any class' and the other criteria is the selection of features that have a higher hedge value 'for every class'. Since, any feature cannot be selective for every class; therefore, a selective function should be described from the hedge values of any feature 'for every class'.

Linguistic hedge (power) values of each feature is determined for each cluster using:

$$p_{ij}(x) = \frac{\prod_{i=1}^{N} \mu_{A_i^j}(x_i)}{\sum_{j=1}^{D} \prod_{i=1}^{N} \mu_{A_i^j}(x_i)} \tag{1}$$

where $p_{ij}(x)$ denotes power value for ith sample and jth feature; N represents total number of samples and D is the total number of features. Power values of features depend on degree of membership functions. Mathematically, power value of feature j is the probability of occurrence of feature j for each cluster. Linguistic Hedge value can be any value in the range $0 \leq p_{ij} \leq 1$. Equation (1) is derived using Bayes' theorem and total probability theorem. Linguistic Hedge Value of features are then aggregated using sum aggregation reasoning rule (SARR) to get a combined contribution of the LH value of features to a particular class. SARR is selected because MIN or MAX reasoning rule performs unsatisfactorily for the problems with overlapping classes, where different features reserve valuable information regarding the class belongingness of a pattern. Each of these features contributes significantly to the desired class and thus the combined effect should be considered to represent the desired class properly. Therefore, a selective function can be described as follow:

$$p_j = \sum_{i=1}^{K} p_{ij} \tag{2}$$

where p_j is the selection value of the jth feature and K is number of classes. Selection value can be any value in the range $0 \leq p_j \leq K$.

1.7. Neuro-Fuzzy Classifier based on SCG algorithm

Neuro-fuzzy classifier (NFC) with Linguistic hedges is based on fuzzy rules. Linguistic hedges are applied to the fuzzy sets of rules, and are adapted by Scaled Conjugate Gradient (SCG) algorithm. In this technique, some distinctive features are emphasized by power values, and some irrelevant features are damped with power values. The effects of power values for particular feature are generally different for different classes. The using of linguistic hedges increases the recognition rates.

NFC partitions the feature space with D input features $\{x_1, x_2, \ldots, x_D\}$ into multiple fuzzy subspaces by fuzzy if-then rules which can be represented by a network structure. A fuzzy classification rule that has D input features $\{x_1, x_2, \ldots, x_D\}$ and one output y is defined with LHs as:

IF x_1 is A_1 with p_1 hedge AND x_2 is A_2 with p_2 hedge ...

... AND x_D is A_D with p_D hedge

THEN y is C_k class.

where $A_1, A_2, \ldots, A_D$ denote linguistic terms that are defined on feature space $\{x_1, x_2, \ldots, x_D\}$; $p_1, p_2, \ldots, p_D$ denote linguistic hedges, respectively; C_k denotes the class label of the output y. Figure 2 shows an architecture of NFC with D selected features $\{x_1, x_2, \ldots, x_D\}$ and three classes $\{C_1, C_2, C_3\}$. NFC consists of input layer, fuzzy membership layer, fuzzification layer, defuzzification layer, normalization layer, and output layer. Each input is defined with three linguistic terms, thus this NFC has 3 fuzzy rules if one cluster is formed for each class. That is, if total number of classes is K and if C numbers of clusters are formed for each class; then total K x C numbers of fuzzy rules are generated. This technique uses zero-order Sugeno fuzzy model for defuzzification. The crisp outputs of fuzzy rules are determined by weighted average operator.

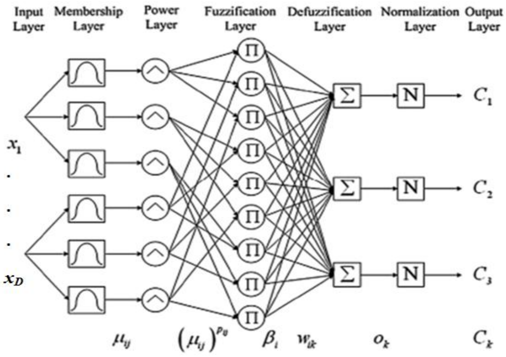

Fig.2. Architecture of Neuro-fuzzy classifier

In this classifier, the nodes in the same layer have the same functions. The functions and properties of each layer are given as follows.

Fuzzy membership layer

The membership grade of each input to specified fuzzy region is measured in membership layer, where generalized bell-shape, Gaussian, triangle, and trapezoidal functions can be used as membership function (MF). In the membership layer, a Gaussian function is adopted, as MF because:

- Continuity of Gaussian functions which is usually required for most of conventional gradient-based, either first- or second-order, optimization techniques.
- It results in smooth partial derivatives of its parameters.
- It has fewer parameters (center and width of Gaussian function).
- A multidimensional Gaussian MF can be decomposed into multiple 1-D Gaussian membership functions for the corresponding number of input variables.

(Hence, separate procedure is not required for partitioning input output space.)

Gaussian MF is defined as follows:

$$\mu_{ij}(x_{sj}) = exp\left[-\frac{1}{2}\left(\frac{x_{sj}-c_{ij}}{\sigma_{ij}}\right)^2\right] \tag{3}$$

where $\mu_{ij}(x_{sj})$ represents the membership grade of ith rule and jth feature. x_{sj} is input data point belonging the sth sample and the jth feature of input matrix $X\{x \in R^{N \times D}\}$. Here, N is the total number of samples and D is the total number of features. Parameters c_{ij} and σ_{ij} represents center and width of Gaussian function, respectively.

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

Power layer

In this layer, the secondary meanings of fuzzy sets are calculated with their FHs:

$$\alpha_{ijs} = [\mu_{ij}(x_{sj})]^{P_{ij}} \tag{4}$$

where α_{ijs} denotes the modified membership grades of $\mu_{ij}(x_{sj})$; P_{ij} denotes the FH value of the ith rule and the jth feature.

Fuzzification layer

Fuzzification layer generates a signal corresponding to the degree of fulfillment of the fuzzy rule for x_s sample. This degree of fulfillment is also called as the firing strength of ith rule. Thus, firing strength β_{is} of the ith rule for x_s sample is as follow:

$$\beta_{is} = \prod_{j=1}^{D} \alpha_{ijs} \tag{5}$$

where D represents the number of features.

Defuzzification layer

Defuzzification is the conversion of a fuzzy quantity to a crisp quantity. The weighted output for the sth sample that belongs to kth class is calculated in defuzzification layer. Each rule is associated with every class according to their weights. However, if a rule controls a particular class region, the weight between this rule output and the specific class is to be greater than the other class weights. Otherwise, the class weights are fairly small. Thus, the weighted output for the sth sample that belongs to kth class is calculated using SARR as follows:

$$O_{sk} = \sum_{i=1}^{U} \beta_{is} W_{ik} \tag{6}$$

where W_{ik} represents the degree of belongingness to the kth class which is controlled with the ith rule. O_{sk} is the weighted output for the sth sample that belongs to the kth class, and U is the total number of rules.

Normalization layer

Sometimes the summation of weights may be greater than 1. Therefore, the outputs of the network should be normalized:

$$h_{sk} = \frac{O_{sk}}{\sum_{i=1}^{K} O_{si}} = \frac{O_{sk}}{\delta_s} \quad \text{where} \quad \delta_s = \sum_{i=1}^{K} O_{si} \tag{7}$$

where h_{sk} represents the normalized degree of the sth sample that belongs to the kth class; and K is the total number of classes. After then, the class label C_s of sth sample is determined by the maximum h_{sk} value using MAX operation.

$$C_s = max\{h_{sk}\} \quad \text{where} \quad k = 1, 2, ..., K \tag{8}$$

1.8. Optimization of weight in NFC

The antecedent parameters of the network $\{c, \sigma, p\}$ could be adapted by any optimization method. In this system, Scaled Conjugate Gradient (SCG) algorithm is used to adapt the network parameters. The SCG is a second order supervised training and derivative based method. It determines the second order derivatives of parameters from their first-order derivatives. In SCG algorithm, negative gradient direction is used to approximate minimum error. In this optimization method, step size is not constant. If error between predicted class and target class is large than step size is increased and when the error approaches to its minimum value;

step size is decreased. Hence, this method is called as Scaled Conjugate Gradient. Hence, this algorithm reduces number of operations in each iteration.

The SCG has a super linear convergence rate, which is two times faster than that of the backpropagation algorithm. The last parameter W_{ik} can also be adapted with the SCG method. However, in the training, W_{ik} can be greater than 1. In these cases, the meanings of weights could be lost among the same-class clusters. Hence, either W_{ik} should be constrained or W_{ik} is determined from the ratio of the number of kth class samples in the ith fuzzy rule region respect to the total number of kth class samples. However, W_{ik} must be determined in the every iteration of optimization method. The weight parameter W_{ik} is assumed as cluster weight. According to the definition explained above, the weight between the ith fuzzy rule and the kth class is given as follows:

$$W_{ik} = \frac{S_i}{S_k} \tag{9}$$

where S_i represents the number of kth class samples that belongs to the ith fuzzy rule region, and S_k denotes the number of all kth class samples. When the fuzzy classification rules are constructed as a network, these parameters can be adapted with neural networks. As a result, superior properties of fuzzy classification systems and neural networks are integrated in this system.

The combined system is named as neuro-fuzzy classifier which is a system based on an adaptive network having multiple inputs multiple outputs.

Computing cost function using least square estimate

The cost function that is used in the SCG algorithm is determined from the least mean square error. Least mean square error is defined as the minimum value of mean square of the difference between target and the calculated class value.

According to the above definition, the cost function E is as follows:

$$\begin{aligned} E &= \frac{1}{N}\sum_{s=1}^{N} E_s, \\ \text{where} \quad E_s &= \frac{1}{2}\sum_{k=1}^{K}(t_{sk} - h_{sk})^2 \end{aligned} \tag{10}$$

where N represents the number of samples; t_{sk} and h_{sk} are target and calculated values of the sth sample belonging to the kth class, respectively. The target value t_{sk} ; which is otherwise 0; is set to 1, if the sth sample belongs to the kth class. For example, let the sth sample belong to the kth class, such that:

$$\begin{aligned} t_s &= [O_{k-1} \quad 1 \quad O_{K-k}]_K \\ \text{where} \quad O_{k-1} &= [0 \quad 0 \quad \ldots \quad 0]_{1\times(k-1)}. \end{aligned} \tag{11}$$

The partial derivative of E with respect to c_{ij} can be calculated using chain rule:

$$\frac{\delta E}{\delta c_{ij}} = \sum_{s=1}^{N} \frac{\delta E}{\delta E_s}\left(\sum_{k=1}^{K} \frac{\delta E_s}{\delta h_{sk}} \cdot \frac{\delta h_{sk}}{\delta O_{sk}} \cdot \frac{\delta O_{sk}}{\delta \beta_{is}} \cdot \frac{\delta \beta_{is}}{\delta \alpha_{ijs}} \cdot \frac{\delta \alpha_{ijs}}{\delta \mu_{ijs}} \cdot \frac{\delta \mu_{ijs}}{\delta c_{ij}}\right) \tag{12}$$

The partial derivatives that are given in equation (6) can be clearly defined as follows:

$$
\begin{aligned}
\frac{\delta E}{\delta E_s} &= \frac{1}{N},\\
\frac{\delta E_s}{\delta h_{sk}} &= h_{sk} - t_{sk},\\
\frac{\delta h_{sk}}{\delta O_{sk}} &= \frac{1-h_{sk}}{\delta_s},\\
\frac{\delta O_{sk}}{\delta \beta_{is}} &= W_{ik},\\
\frac{\delta \beta_{is}}{\delta \alpha_{ijs}} &= \frac{\beta_{is}}{\alpha_{ijs}},\\
\frac{\delta \alpha_{ijs}}{\delta \mu_{ijs}} &= \frac{p_{ij}}{\mu_{ijs}}.\alpha_{ijs},\\
\frac{\delta \mu_{ijs}}{\delta c_{ij}} &= \mu_{ijs}.\frac{x_{sj}-c_{ij}}{\sigma_{ij}^2}
\end{aligned}
\tag{13}
$$

Similarly, the partial derivatives of E respect to σ_{ij} and p_{ij} can also be defined respectively. The Neuro-fuzzy classifier based on linguistic hedges is trained with the SCG optimization algorithm using the partial derivatives of the cost function E with respect to the above parameters.

III. RESULTS

This section briefs about the results obtained after implementation of neuro-fuzzy classifier for thyroid disease diagnosis.

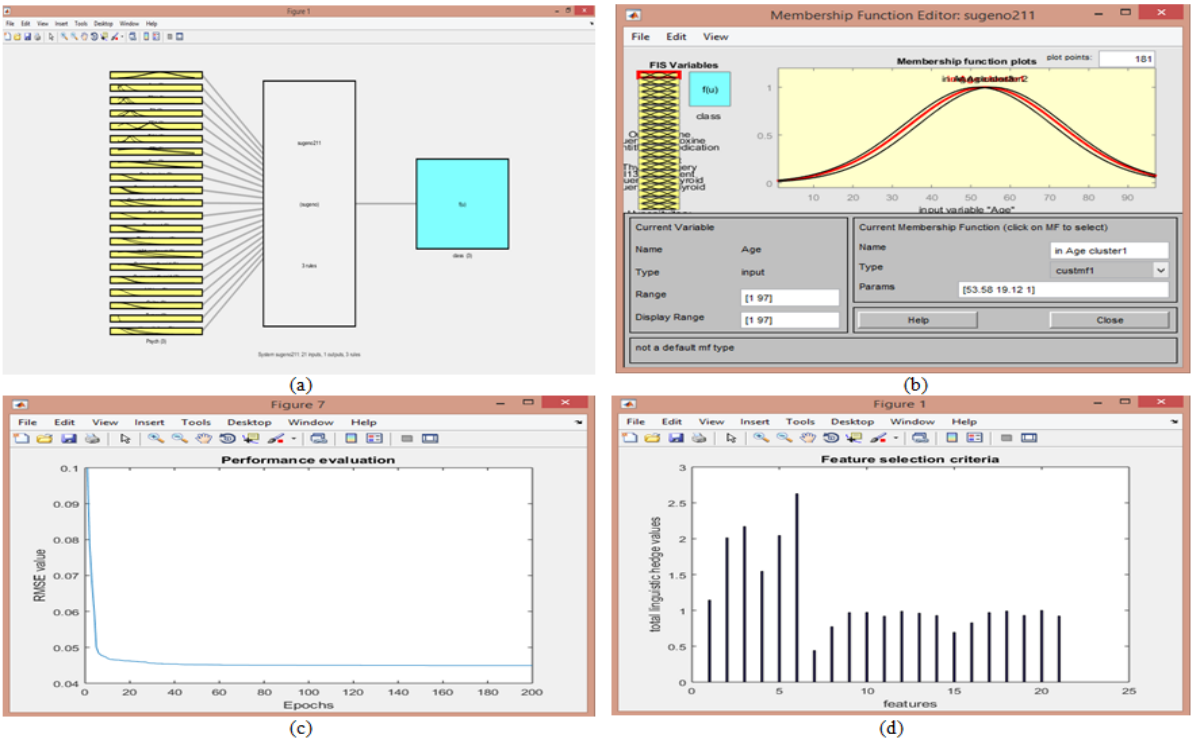

Fig.3. Results obtained after performing feature selection using linguistic hedges.

Fig.3 (a) shows model representation of neuro-fuzzy system to which 21 attributes are applied as inputs and has 1 output variable class which can take value as 1, 2 or 3 according to type of diagnosis i.e. normal, hyperthyroid or hypothyroid. This system uses Sugeno type of Fuzzy Interference System.

The input features are fuzzified using Guassian MF according to their class using K-means clustering as shown in Fig.3 (b). The parameters of this MF i.e. center, width, degree of MF are obtained using SCG training algorithm.

Mean Square Error is used as the cost function for training the MF. Fig.3 (c) shows plot of RMSE verses number of epochs (iterations). Minimum mean square error reached is 0.0449.

Fig.3 (d) shows bar graph showing selection value for each feature. Features are indicated on horizontal axis using feature indices from 1 through 21. Higher the linguistic hedge value of any particular feature; greater is the significance of that feature.

Features 7 through 21 have similar linguistic hedge value around 0.9. Hence, the feature selection criteria used is such that all the features having selection power values greater than or equal to 1 are selected for classification of thyroid disease. Then only first six features having power values greater than 1 are selected and all others are rejected. The six selected features are FTI, T3, T4U, TSH, TT4, and Age.

Fig.4 (a) shows model representation of neuro-fuzzy system to which 6 selected features are applied as inputs and has 1 output variable class which can take value as 1, 2 or 3 according to type of diagnosis i.e. normal, hyperthyroid or hypothyroid. This system uses Sugeno type of Fuzzy Interference System.

To overcome problems of overlapping ranges for different class for one particular feature; the number of clusters for each class can be increased. Hence, for this proposed system two clusters are defined for each class. Since, there are three classes, total number of clusters in this system is six. Hence, this system will have 6 fuzzy rules defined for classification.

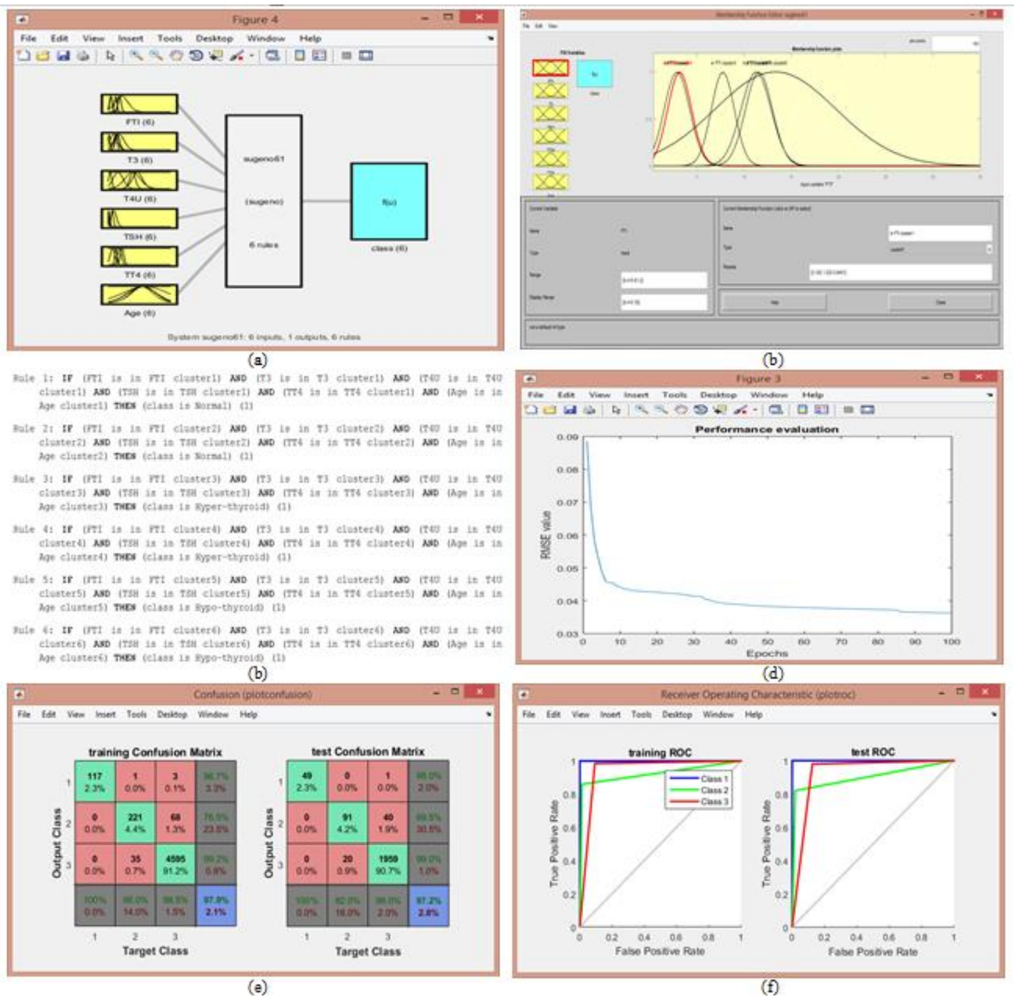

Fig.4. Results obtained after performing classification using neuro-fuzzy classifier.

The selected input features are fuzzified using Guassian MF according to their class using K-means clustering as shown in Fig.4 (b). Here, for each class two clusters are defined. The parameters of this MF i.e. center, width, degree of MF are adapted using SCG training algorithm.

Fig.4 (c) shows six classification rules defined using t-norms. The proposed system is based on sugeno-type of fuzzy interference system.

Mean Square Error is used as the cost function for training the MF during classification. Fig.4 (d) shows plot of RMSE verses number of epochs (iterations). Minimum mean square error reached is 0.0362.

Confusion matrices illustrating classification accuracies of training and test sets are shown in Fig.4 (e). The correctly classified samples are shown diagonally in green boxes. All other misclassified samples are shown in red boxes. The classification accuracies for training set and test sets are 97.9 and 97.2 respectively.

ROI plots obtained for training and test sets applied to Neuro-fuzzy Classifier are shown in Fig.4 (f).

ROC graphs are another way besides confusion matrices to examine the performance of classifiers. A ROC graph is a plot with the false positive rate on the *X* axis and the true positive rate on the *Y* axis. The point (0, 1) is the perfect classifier: it classifies all positive cases and negative cases correctly. It is (0, 1) because the false positive rate is 0 (none), and the true positive rate is 1 (all). The point (0, 0) represents a classifier that predicts all cases to be negative, while the point (1, 1) corresponds to a classifier that predicts every case to be positive. Point (1, 0) is the classifier that is incorrect for all classifications. In many cases, a classifier has a parameter that can be adjusted to increase *TP* at the cost of an increased *FP* or decrease *FP* at the cost of a decrease in *TP*. Each parameter setting provides a (*FP*, *TP*) pair and a series of such pairs can be used to plot an ROC curve. A non-parametric classifier is represented by a single ROC point for a particular data set, corresponding to its (*FP*, *TP*) pair.

Fig.4 (f) shows ROC graph with three ROC curves labeled C1, C2 and C3.

Features of ROC Graphs

- An ROC curve or point is independent of class distribution or error costs.
- An ROC graph encapsulates all information contained in the confusion matrix, since *FN* is the complement of *TP* and *TN* is the complement of *FP*.
- ROC curves provide a visual tool for examining the tradeoff between the ability of a classifier to correctly identify positive cases and the number of negative cases that are incorrectly classified.

IV. CONCLUSION AND FUTURE WORK

In this paper, neuro-fuzzy classifier using linguistic hedges is proposed for thyroid disease diagnosis which is trained using Scaled Conjugate Gradient algorithm. The results obtained from the model are compared and analyzed to enhance the use of neuro-fuzzy classifier in the field of medical diagnosis.

Further research could comprise of training the classifier using training algorithms other than SCG algorithm and develop best algorithm for fast and accurate disease diagnosis on the basis of its accuracy, time taken to build the model. Future work may also include varying training dataset size, taking additional number of inputs and output targets for classification, to enhance the utility of the neuro-fuzzy systems in the field of diagnosis of medical diseases. In this paper, the neuro-fuzzy classifier was trained to diagnose thyroid diseases, other major diseases could also be predicted by training the classifier with appropriate data and enhance the medical diagnosis system.

V. ACKNOWLEDGEMENTS

This project has endured a long journey from concept to structured framework and then to its final implementation. I would like to thank my internal guide, Ms. Amita Dessai, Assistant Professor, ETC Department, for allowing me to carry out this project under her supervision and for her valuable suggestions, encouragement and constructive criticism. I would also like to express my gratitude to Dr. H. G. Virani, Professor and Head of Department, Electronics and Telecommunication Engineering, and Dr. V. N. Shet, Principal, Goa College of Engineering for allowing me to use the college facilities.

REFERENCES

[1] Jamil Ahmed, and M. Abdul Rehman Soomrani, "TDTD: Thyroid Disease Type Diagnostics." 2016 International Conference on Intelligent Systems Engineering (ICISE), pp. 44-50, IEEE, 2016, DOI: 10.1109/INTELSE.2016.7475160.

[2] Prerana, Parveen Sehgal, and Khushboo Taneja, "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network." International Journal of Research in Management, Science & Technology (E-ISSN: 2321-3264) Vol 3, No. 2, April 2015.

[3] I Md. Dendi Maysanjaya, Hanung Adi Nugroho and Noor Akhmad Setiawan, "A Comparison of Classification Methods on Diagnosis of Thyroid Diseases.", International Seminar on Intelligent Technology and Its Applications (ISITIA), pp. 89-92 IEEE, 2015, DOI: 10.1109/ISITIA.2015.7219959.

[4] Shanu Shroff, Siddhi Pise, Pratiksha Chalekar and Suja S. Panicker, "Thyroid Disease Diagnosis: A survey.", IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO), pp. 1-6, 2015, DOI: 10.1109/ISCO.2015.7282384.

[5] Gurmeet Kaur and Er. Brahmaleen Kaur Sidhu, "Proposing Efficient Neural Network Training Model for Thyroid Disease Diagnosis.", International Journal For Technological Research In Engineering Volume 1, Issue 11, ISSN (Online): 2347 - 4718, pp. 1383-1386, July-2014.

[6] Jasdeep Singh Bhalla and Anmol Aggarwal, "A Novel Method for Medical Disease Diagnosis Using Artificial Neural Networks Based on Backpropagation Algorithm.", Conuence 2013: The Next Generation Information Technology Summit (4th International Conference). IET, pp. 55-61, September 26, 2013, IEEE, June,2014, DOI: 10.1049/cp.2013.2293.

[7] Modjtaba Rouhani and Kamran Mansouri,\Comparison of several ANN architectures on the Thyroid diseases grades diagnosis.", International Association of Computer Science and Information Technology – Spring Conference, 2009, IACSITSC'09, pp. 526-528, IEEE, April 17, 2009, DOI: 10.1109/IACSIT-SC.2009.24.

[8] Ahmad Taher Azar and Shaimaa Ahmed El-Said, "Superior neuro-fuzzy classification systems.", Neural Computing and Applications (2012), 23.1 : pp. 55-72, November 2012, DOI: 10.1007/s00521-012-1231-8.

[9] Bayram Cetili and Atalay Barkana, "Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training.", Soft computing (2010), 14.4 : pp. 365-378, DOI: 10.1007/s00500-009-0410-8.

[10] Bayram Cetisli, "Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1.", Expert Systems with Applications (2010), 37.8 : pp. 6093-6101, DOI: 10.1016/j.eswa.2010.02.108.

[11] Bayram Cetisli, "The effect of linguistic hedges on feature selection: Part 2.", Expert Systems with Applications (2010), 37.8 : pp. 6102-6108, DOI: 10.1016/j.eswa.2010.02.115.

[12] Ahmad Taher Azar and Aboul Ella Hassanien, "Dimensionality reduction of medical big data using neural-fuzzy classier.", Soft Computing (2014), DOI: 10.1007/s00500-014-1327-4.

[13] Thyroid Data Set from UCI Repository [Online] Available: http://sci2s.ugr.es/keel/dataset.php?cod=67