Kuwait J. Sci. **42 (2)** pp. 189-205, 2015

Optimized construction of various classification models for the diagnosis of thyroid problems in human beings

RAJKUMAR NALLAMUTH* AND JACANATHAN PALANICHAMY**

**Associate Professor, Department of Computer Applications, PSNA College of Engineering and Technology, India, rknpsna@gmail.com*
***Professor, Department of Computer Applications, PSNA College of Engineering and Technology, India, jaganathodc@gmail.com*
**Corresponding author: rknpsna@gmail.com*

ABSTRACT

Thyroid disorder is a major public health problem. Early detection of thyroid disorder is an increasingly important area in the field of medical diagnosis, pattern recognition, machine learning and data mining. Thyroid disorder, either over production (hyperthyroidism) or less production (hypothyroidism) results in imbalanced state of thyroid hormone stimulation in human beings. So, controlling this disorder has become a central issue in healthcare and needs great attention. This research critically examines different classification models constructed using a novel mathematical ranked improved F-score ordering (RIFO) applied to thyroid dataset taken from machine learning repository, University of California, Irvine. A total of nine possible and effective feature subsets have been constructed and each subset is tested with three most benchmarked algorithms namely C4.5, multilayer perceptron (MLP) and radial basis function network (RBFN) using tenfold cross-validation and various training-test partitions. The obtained results show diverse conclusions, but one with interesting and highest accuracy has been presented. From the results, it is observed that MLP has emerged with an outstanding performance of 98.15%, which is greater than all earlier research. The dataset has 3 classes, 5 features and 215 records (hypo=30, hyper=35, normal=150).

Keywords and phrases: Thyroid disorder; ranked improved F-score ordering; C4.5; MLP; RBFN.

INTRODUCTION

Millions of people in the world have thyroid disorders especially women. Most of them have undiagnosed thyroid diseases, which becomes a major concern. An abnormality or imbalance in production of thyroid hormones can cause a host of problems, from fatigue and depression to weight gain leading to thyroid problems. The thyroid gland is the largest gland among the endocrine glands of the human body and is placed in the middle of the lower neck, below the larynx and above the clavicles. It produces two active thyroid hormones, levothyroxine (T4) and triiodothyronine (T3). These hormones help in regulating the body's metabolism (Zhang & Berardi, 1998). The most common problems are hypothyroidism, hyperthyroidism and thyroid nodules. Hypothyroidism is a disorder with multiple causes in which the thyroid fails to secrete an adequate amount of thyroid hormone. Hypothyroidism can also be associated with pregnancy. Hyperthyroidism refers to excess secretion of thyroid hormones, which results in accelerated metabolism in peripheral tissues. Grave's disease is the cause of irregular production of thyroid hormone in human beings. It is a condition that the proteins consistently instruct the thyroid to produce more thyroid hormone (Ozyılmaz & Yıldırım, 2002). A thyroid nodule is an abnormal growth of thyroid cells within the thyroid gland. The majority of thyroid nodules are benign (non cancerous), only a small portion of thyroid nodules have thyroid cancer. The disease can be diagnosed through proper examination and focused interpretation of the thyroid data. Preservation of joint relationships between variables/data is vital (Niloofar *et al*., 2013) and the classification of thyroid functioning in different classes of thyroid patients is an important problem(Hoshi, 2005). Earlier many methods, like pattern recognition, fuzzy techniques, artificial immune recognition system, neural networks, etc., have been used in categorizing thyroid patients (Delen *et al*., 2005).

Features are an important catalyst in data mining classification systems. Data mining techniques are much useful in feature selection process. It is finding increasing applications in expertise orientation and the development of applications for data mining techniques is a problem-oriented domain (Shu-Hsien *et al*., 2012). Commonly, when a feature selection algorithm is applied, a single feature subset is selected for all the classes, but this subset could be inadequate for some classes (Barbara *et al*., 2011). Also, a feature selection process motivated by medical knowledge is important (Nahar *et al*., 2013). Therefore, constructing various subsets of features from the available feature set could substantially help in improved diagnosis of a disease.

In this study, different models have been constructed by ranked ordering of their calculated improved F-Score values. The abnormalities of thyroid disease were classified using three classifiers namely C4.5, MLP and RBFN. Multilayer perceptron (MLP) and radial basis function (RBF) are two important neural network structures, which are good in pattern classification problems. They are robust classifiers with the ability to generalize from imprecise input data. The general difference between MLP and RBF is that RBF is a localist type of learning, which is responsive only to a limited section of input space. On the other hand, MLP is more distributed approach. The output of a MLP is produced by linear combinations of the outputs of hidden layer nodes, in which every neuron maps a weighted average of the inputs through a sigmoid function. In a hidden layer, RBF network hidden nodes map distances between input vectors and center vectors to outputs through a nonlinear kernel or radial function (Yilmaz & Kaynar, 2011). The multiclass classification is one of the fundamental tasks in machine learning. Many algorithms have been proposed for multiclass classification based on the artificial neural network (ANN). The mostly adopted network topology is radial basis function neural network (RBFNN) due to a number of advantages compared with other types of ANNs, such as better prediction accuracy, simpler network structure, and faster learning process (Mitchell, 2003). Recently, many variants of RBFNNs were invented to solve multiclass classification problems (Fu & Wang, 2003).

The remaining part of the paper comprises of the related work on thyroid disease diagnosis, construction of classification models using RIFO, explanation of the classifier algorithms such as C4.5, MLP and RBFN, experimental results, interpretation of proposed procedure, results, conclusion and further research.

RELATED WORK

In recent years, there has been an increased amount of literature on thyroid disease classification. There have been several studies reported focusing on thyroid disease diagnosis (Pasi, 2004). Serpen *et al*. (1997) obtained different classification accuracies using MLP (36.74%), LVQ (81.86%), RBF (72.09%) and PPFNN (78.14%) (Serpen *et al*., 1997). MLP with back propagation produced 86.33%, RBF produced 79.08% and CSFNN produced 91.14% for Ozyilmaz and Yıldırım (Ozyılmaz & Yıldırım, 2002). Pasi obtained the following accuracies with LDA (81.34%), C4.5-1 (93.26%), C4.5-2 (92.81%), C4.5-3 (92.94), MLP (96.24%) and DIMLP (94.86%). Polat obtained 85.00% and 81.00% of accuracies using AIRS with Fuzzy weighted pre-processing and AIRS respectively (Polat *et al*., 2007). Keles & Keles, (2008) obtained 95.33% with Neuro Fuzzy Classification (ESTDD with NEFCLASS-J) (Keles *et al*., 2008). Feyzullah Termurtas, (2009) produced different classification accuracies using MLNN with LM (92.96%) for 3-fold cross-validation, PNN (94.43%) for 3-fold cross-validation, LVQ (89.79%) for 3-fold cross-validation, MLNN with LM (93.19%) for 10-fold cross-validation, PNN (94.81%) for 10-fold cross-validation, LVQ (90.05%) for 10-fold cross-validation. Doganteken *et al*. (2011) obtained 91.86% of classification accuracy using GDA-WSVM (Dogantekin *et al*., 2011). Jaganathan and Rajkumar obtained an accuracy of 93.49% with multilayer perceptron (Jaganathan & Rajkumar, 2012).

CONSTRUCTION OF CLASSIFICATION MODELS

It is natural to wish to examine data as they accumulate during the course of a long-term clinical trial (Christopher & Bruce, 2013). Classification is one of the most intelligent techniques for data analysis that can be used to extract models in machine learning, statistics and data mining. Likewise, feature selection is a technique for reducing dimensionalities in the field of medical diagnosis, machine learning and data mining. In general, the aim of feature selection is to determine the most relevant combination of features, which will strongly predict the target class in a specific problem. The main objective of this research is to investigate all possible classification models that can be constructed for the thyroid dataset. In the current study, in classifying thyroid diseases, the improved F-score value of each of the features constituting all the three classes has been computed. The computed scores are sequentially ranked from their low score to high score and vice versa. After ranking, the models were constructed based on the number of features for each model. Initially, the model started with the feature having lowest value and the total numbers of features in the models were increased by the next highest value each time and vice versa. A total of nine models from both ordering have been constructed as shown in the Table 4. Hence, this study, has applied a ranked improved F-score ordering for constructing classification models. The noteworthy contribution of the work is that the results obtained generated a best model which enhanced the classification accuracy.

Computation of improved f-score

Earlier studies reveal that an intelligent technique is a must for measuring the discrimination when there are more than two classes. Improved F-score is one such technique to evaluate the databases with multi classes like thyroid dataset which has three classes. The improved F-score value of a feature is computed using equation (1).

$$F_i = \frac{\sum_{j}^{l} =_1 \left(\bar{x}_i^{(j)} - \bar{x}_i\right)^2}{\sum_{j}^{l} = \frac{1}{1_{n_j-1}} \sum_{k=1}^{n_j} \left(x_{k,j}^{(j)} - \bar{x}_i^{(j)}\right)^2} \tag{1}$$

where l denotes the number of classes in a dataset; $\bar{x}_i, \bar{x}_i^{(j)}$ are the averages of the i^{th} feature of the whole dataset and the jth dataset respectively; $x_{k,i}^{(j)}$ is the i^{th} feature of the k^{th} instance in the j^{th} dataset. The numerator denotes the discrimination between each dataset where as the denominator indicates the one within each of dataset (Xie & Wang, 2011).

CLASSIFICATION

Multilayer perceptron (MLP)

A multilayer perceptron is a feed forward neural network and it consists of multiple layers, where each layer is composed of nodes and each node connects to every node in subsequent layers. Multilayer perceptron is composed of a minimum of three layers such as input layer, one or more hidden layers and an output layer. The input layer gives the input to subsequent layers. The input nodes have linear activation functions and no thresholds. All hidden unit nodes and output nodes have thresholds associated with them, in addition to the weights. The hidden nodes have nonlinear activation functions and the outputs have linear activation functions.

Multilayer perceptron is a network of simple neurons. The perceptions produce a single output from multiple inputs based on its input weights with nonlinear activation function. This activation function can mathematically be written as

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(W^T X + b) \tag{2}$$

where W indicates the weights, X denotes the inputs, b is the bias and φ is the activation function (Du *et al*., 2006).

C4.5

C4.5 is a statistical classifier used to generate a decision tree developed by Ross Quinlan (Quinlan, 1996). It builds decision trees from a set of training data using the concept of information entropy. The training data is a set $S = s_1, s_2,...$ of already classified samples. Each sample $s_i = x_1, x_2,...$ is a vector where $x_1, x_2,...$ represent features of the sample. The set of training data is augmented with a vector $C = c_1, c_2,...,$ where $c_1, c_2,...$ represent the class to which each sample belongs. At each node, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criteria are the normalized information gain or difference in entropy that results from choosing an attribute for splitting the data. The attribute with the greater normalized information gain is chosen to make the decision. The C4.5 algorithm then iterates on the smaller sub lists (Quinlan, 1996).

Radial basis function network

Neural networks are powerful computational methods that have resulted from work in the area of neurophysiology and are being applied to an increasing range of medical problems (Paliwal & Kumar, 2009). Radial basis function networks is an artificial neural network that uses radial basis functions as activation functions that are widely used in function approximation, time series prediction and control. RBF Networks have an input layer, a hidden layer with a non-linear RBF activation and an output layer.

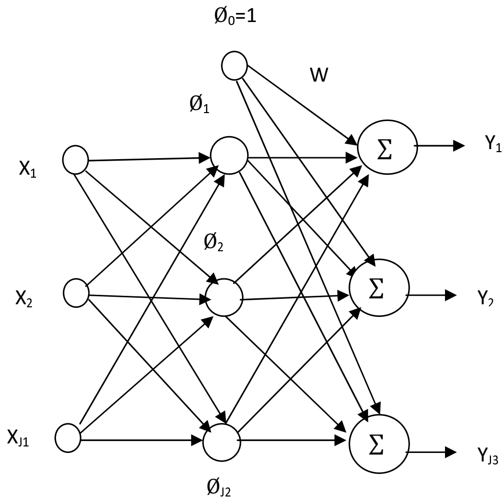

Fig. 1 Architecture of RBFN

Input layer consists of source nodes that connect network to its environment. Hidden layer applies a non-linear transformation from the input space to the hidden space, which is high dimensional. The output of the network is linear combination of the outputs from radial basis functions. The transformation from the input to hidden layer is nonlinear, but transformation from hidden layer to output space is linear.

In Figure 1. For a given input vector x, the output of the network is given by

$$y_i(x) = \sum_{k=1}^{j_2} w_{ki}\phi(||x - c_k||) \tag{3}$$

for i=1,……,J_3 where $y_i(x)$ is the i^th^ output of the RBFN, w_{ki} is the connection weight from the k^th^ hidden unit to the i^th^ output unit, c_k is the centre of the k^th^ hidden unit, and ||.|| is representing Euclidean norm. The RBF Ø(.) is typically selected as the Gaussian function. Each node in the hidden layer uses an RBF, denoted by Ø(r). The biases of the output layer neurons are modeled by an additional neuron in the hidden layer, which has a constant activation function $Ø_0$(r) (Du & Swamy, 2006).

EXPERIMENTAL OBSERVATION

Thyroid gland dataset

The thyroid gland dataset used in the study is taken from UCI machine learning repository. They have been collected by Danny Coomans (1992) at the University of James Cook (Coomans *et al*.,). The dataset consists of 3 classes, which are normal, hyper, hypo function of the thyroid gland; 215 samples and each sample has 5 attributes. These 5 attributes are described in Table 1. Among the 215 samples, 150 samples belong to normal- function (class-1), 35 samples belong to hyper-function (class-2) and 30 samples belong to hypo-function of the thyroid gland dataset.

Table 1. Attribute description of thyroid gland dataset

Attribute number	Attribute description
F1	T3-resin uptake test (A percentage)
F2	Total serum thyroxin as measured by the isotopic displacement method
F3	Total serum triiodothyronine as measured by radioimmuno assay
F4	Basal thyroid-stimulating hormone(TSH) as measured by radioimmuno assay
F5	Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotrophin-releasing hormone as compared to the basal value

Table 2. Results obtained from previous research

Author	Technique	Performance/ Accuracy (%)
Ozyilmaz and ildirim (2002)	RBF	79.08
Ozyilmaz and ildirim (2002)	CSFNN	91.14
Serpen et al. (1997)	MLP	36.74
Serpen et al. (1997)	LVQ	81.86
Serpen et al. (1997)	RBF	72.09
Serpen et al. (1997)	PPFNN	78.14
Pasi (2004)	LDA	81.34
Pasi (2004)	C4.5-1	93.26
Pasi (2004)	C4.5-2	92.81
Pasi (2004)	C4.5-3	92.94
Pasi (2004)	MLP	96.24
Pasi (2004)	DIMLP	94.86
Keles and Keles (2008)	ESTDD with NEFCLASS-J	95.33
Polat *et al.* (2007)	AIRS	81.00
Polat *et al.* (2007)	AIRS with fuzzy weighed reprocessing	85.00
Dogantekin *et al*. (2010)	ADSTG	97.67
Dogantekin *et al*. (2011)	GDA-WSVM	91.86
Jaganathan and Rajkumar (2012)	Improved F-score with MLP	93.49
This study	RIFO with MLP	98.15

The experiments were carried out using a Java application developed by the University of Waikato in New Zealand, Weka software. Multilayer perceptron is a classifier that uses back propagation to classify records. The following parameters were used in designing MLP network. They are three hidden layers with three nodes, learning rate L = 0.3, Momentum M = 0.2, Epochs = 500, and Sigmoid function. Learning rate is the amount in which the weights are updated. M is the momentum which is applied to the weights during updating. C4.5 is a classifier for generating a pruned or unpruned C4.5 decision tree. The parameters used in running C4.5 were confidence factor C = 0.25 used for pruning and number of instances per leaf M is 2. RBFN is a class that implements normalized Gaussian radial basis function network. It uses the k-means clustering to provide the basis functions and learns either a logistic regression (discrete class problems) or linear regression (numeric class problems) on top of that. Also, symmetric multivariate Gaussians are fit to the data from each cluster. When the class is nominal it uses the given number of clusters per class. It standardizes the numeric attributes to zero mean and unit variance. The parameters used in RBFN are -B 2 -S 1 -R 1.0E-8 -M -1 -W 0.1, where S is random clustering seed to pass on to K-means. M is Maximum number of iterations for the logistic regression to perform. W sets the minimum standard deviation for the clusters. B is the number of clusters for K-means to generate. R is the ridge value for the logistic or linear regression. MLP, C4.5 and RBFN are simple and efficient classifiers based on solid mathematical grounds. MLP has the remarkable ability to train fast and extract patterns from complex data (Jaganathan & Rajkumar, 2012). It is also the most widely used type of neural network in the research community. C4.5 is one of the prominent machine learning algorithms used in modern computing, which produces useful results in classification task. Moreover, these algorithms have been used in thyroid disease diagnosis earlier and have produced better results (Feyzullah Temurtas, 2009). In order to evaluate the efficiency of the method, performance measures like sensitivity, specificity and ROC curves were considered. The measures were compiled by the following units.

$$\text{Sensitivity } (\%) = \frac{TP}{(TP+FN)} \times 100 \tag{3}$$

$$\text{Specificity } (\%) = \frac{TN}{(FP+TN)} \times 100 \tag{4}$$

$$\textit{Classification accuracy } (\%) = \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100 \tag{5}$$

ROC Curve provides trade-off between sensitivity and specificity values.

Table 3. Confusion matrix representation

Actual	**Predicted**	
	Positive	Negative
Positive	TP: True positive	FN: False negative
Negative	FP: False positive	TN: True negative

RESULTS AND DISCUSSION

Most studies in the field of thyroid dataset classification have only focused on feature selection and classification. Though extensive research has been carried out, an adequate study to construct different classification models, which will enhance the classification accuracies have not been covered. In this paper, a serious attempt has been made to construct various classification models using ranked improved f-score ordering (RIFO). Three algorithms namely C4.5, RBFN and Multilayer perceptron have been used to obtain classification accuracies. Table 1 shows the attribute description of thyroid gland dataset. Table 2 shows accuracies of earlier research so far and the highest is found to be 97.67 using ADTSG method (Dogantekin *et al.*, 2010). Table 3 shows confusion matrix representation. Table 4 gives the computed improved F-score values of each feature and all the constructed models based on the proposed technique. The improved F-score values each feature are as follows. F1=0.35, F2=1.15, F3=1.49, F4=1.40, F5=1.26.

Table 4. Models constructed from ranked improved F-score ordering (RIFO)

Features	**Improved F-score value**	**Rank based on improved F-score**	**Models**	**Models based on ranked ordering from lowest -highest improved F-score value**	**Models**	**Models based on ranked ordering from highest-lowest improved F-score value**
F1	0.35	5 (F1)	#1	F1	#6	F3
F2	1.15	4 (F2)	#2	F1,F2	#7	F3,F4
F3	1.49	1 (F3)	#3	F1.F2,F5	#8	F3,F4,F5
F4	1.40	2 (F4)	#4	F1,F2,F5,F4	#9	F3,F4,F5,F2
F5	1.26	3 (F5)	#5			

The features were ranked based on their lowest improved F-score value to highest and the ranked order is found to be F1<F2<F5< F4<F3. Next, the features were ranked based on their highest improved F-score value to lowest. In this case, the ranked order

is found to be F3>F4> F5>F2>F1. Based on the above orderings, the constructed nine different subset models are {F1},{F1,F2},{F1,F2,F5},{F1,F2,F5,F4},{F1,F2,F5,F4 ,F3} for low value to high value and {F3}, {F3,F4},{F3,F4,F5},{F3,F4,F5,F2},{F3,F 4,F5,F2,F1} for high value to low value.

Table 5 is quite revealing in many ways. It shows the performance of the classifiers with 5 different subset models constructed from lowest to highest using ten-fold cross-validations and various training-test partitions. The best observed results are 98.15%, 96.00 % and 96.91% for

Table 5. Classification accuracies for each model using RIFO from lowest to highest

Algorithm	**Various training-test partitions**	**#1**	**#2**	**#3**	**#4**	**#5**
		F1	**F1,F2**	**F1,F2,F5**	**F1,F2,F5,F4**	**F1,F2,F5,F4,F3**
MLP	55-45%	72.16	90.72	94.85	94.85	92.78
	65-35%	76.00	89.33	96.00	94.67	94.66
	75-25%	75.93	88.89	98.15	96.30	90.74
	85-15%	75.00	8750	96.88	96.87	96.87
	Ten-fold	79.53	91.16	93.95	93.95	95.82
C4.5	55-45%	72.16	86.60	89.69	90.72	90.72
	65-35%	76.00	89.33	96.00	94.67	90.67
	75-25%	75.93	87.04	94.44	92.59	90.74
	85-15%	71.87	90.63	93.75	93.75	93.75
	Ten-fold	77.67	88.84	94.42	93.95	93.49
RBFN	55-45%	72.16	92.78	96.91	95.88	95.87
	65-35%	76.00	90.67	94.67	96.00	94.67
	75-25%	77.78	90.74	96.30	96.30	94.44
	85-15%	75.00	90.63	93.75	90.63	90.63
	Ten-fold	79.07	92.56	95.35	94.88	95.35

Multilayer perceptron, C4.5 and RBFN respectively. In this case, what is interesting is the overall best performance is 98.15% for MLP using 75-25% training-test partitions, which is better than the best so far achieved in the literature. Similarly, Table 6 shows the performance of the classifiers with different subset models constructed from high to low, using ten-fold cross-validations and various training-test partitions. Data from this table can be compared with the data in Table 5, which clearly indicate the different accuracies generated by each model under different circumstances. Here, the obtained classification accuracy for MLP is 95.88% using 55- 45% training-test partitions. The accuracy for C4.5 is 93.75% using 85-15% training-test partitions and accuracy for

RBFN is 96.87% using 85-15% training-test partition. Among all, the most striking result to emerge is that MLP has produced 98.15% of classification accuracy with the model #3 consisting of features {F1, F2, F5}. This result is higher than the most recent and all the earlier research results. Table 7 represents the relative individual feature relevance of each feature with target class. The classification accuracies were computed for each feature individually for comparing the relative importance of each feature with the target class using the same benchmarked algorithms MLP, C4.5 and RBFN. The results show that F1 best is = 79.53%, F2 best is = 89.77%, F3 best is = 85.12%, F4 best is = 80.47% and F5 best is = 77.68%. This indicates that individually the features have not been able to accurately classify the instances.

Table 6. Classification accuracies for each model using RIFO from highest to lowest

Algorithm	**Various training-test partitions**	**#6** **F3**	**#7** **F3,F4**	**#8** **F3,F4,F5**	**#9** **F3,F4,F5,F2**
MLP	55-45%	82.47	87.63	88.66	94.85
	65-35%	82.67	88.00	89.33	93.33
	75-25%	81.48	87.00	88.89	92.59
	85-15%	81.25	84.38	87.50	87.50
	Ten-fold	85.11	90.70	90.70	93.95
4.5	55-45%	82.47	87.63	87.63	87.63
	65-35%	82.67	88.00	88.00	88.00
	75-25%	81.48	87.04	87.03	88.89
	85-15%	81.25	84.38	84.38	90.63
	Ten-fold	85.12	91.16	91.16	92.56
RBFN	55-45%	82.47	87.63	88.66	92.78
	65-35%	82.67	88.00	88.00	92.00
	75-25%	81.48	87.04	88.89	92.59
	85-15%	81.25	84.38	84.38	90.63
	Ten-fold	85.12	90.70	91.16	93.49

Table 7. Relative importance each feature with improved F-score

Algorithm	Various training-test partitions	Feature1	Feature 2	Feature 3	Feature 4	Feature 5
MLP	55-45%	72.16	89.69	82.47	76.29	73.20
	65-35%	76.00	86.67	82.67	76.00	74.67
	75-25%	75.92	85.19	81.48	74.07	72.22
	85-15%	75.00	87.50	81.25	65.63	65.63
	Ten fold	77.64	89.76	85.12	80.47	77.64
C4.5	55-45%	72.20	89.69	82.47	76.29	73.20
	65-35%	76.00	86.67	82.67	76.00	74.67
	75-25%	75.92	85.19	81.48	74.07	72.22
	85-15%	71.88	87.50	81.25	65.63	65.63
	Ten fold	79.53	89.77	85.12	80.45	77.67
RBFN	55-45%	72.16	86.60	82.47	76.29	73.20
	65-35%	76.00	82.67	82.67	76.00	74.67
	75-25%	77.77	85.19	81.48	74.07	72.22
	85-15%	75.00	87.50	81.25	65.63	65.63
	Ten-fold	79.07	88.37	85.11	80.46	77.68

F1 Best: 79.53 - F2 Best: 89.77 - F3 Best: 85.12 - F4 Best: 80.47 - F5 Best: 77.68

Table 8. Performance evaluation of #3(F1,F2 &F5)

Methods	75% - 25% training-test partition			Ten-fold Cross-validation		
	Sensitivity	Specificity	AUC Value	Sensitivity	Specificity	AUC Value
MLP	98.15	96.60	0.995	94.00	91.00	0.98
C4.5	94.40	95.50	0.957	94.40	92.10	0.94
RBFN	96.29	96.10	0.982	95.30	94.20	0.97

There is no evidence for any significant improvement of classification performances. But it is strongly evident; the present study has produced higher and more accurate classifications.

Table 8. demonstrates the performance evaluation of the best performing feature subset model #3 (F1,F2,F5) using 75-25% training-test partition and ten-fold cross-validations which produced improved performance of 98.15%. The sensitivity produced by MLP, C4.5, RBFN are 98.15%, 94.40%, 96.29%, respectively using 75-25% training-test partitions and 94.00%, 94.40%, 95.30% respectively using ten-fold cross-validations.

ROC curves obtained from various classifiers and methods:

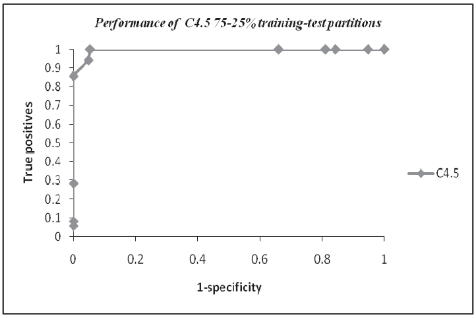

Fig.2. ROC for MLP 75-25% partition

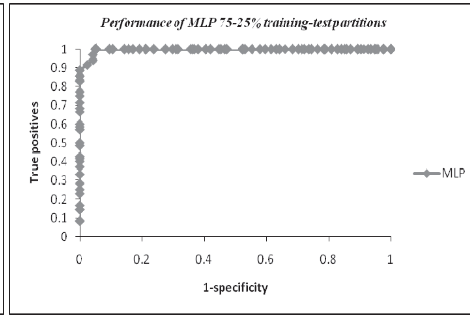

Fig.3. ROC for C4.5 75-25% partition

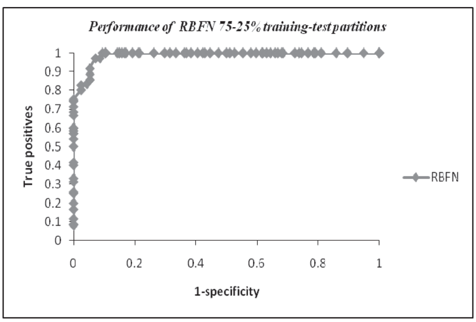

Fig.4 ROC for RBFN 75-25% partition

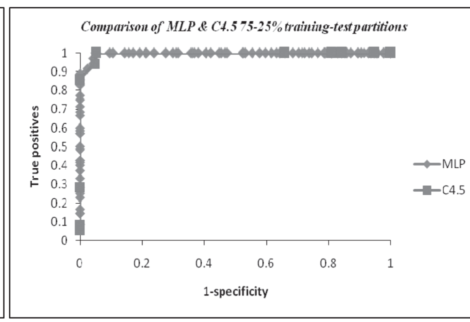

Fig.5 ROC for MLP & C4.5 75-25% partition

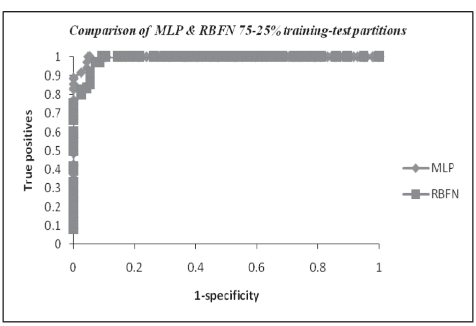

Fig.6. ROC for MLP & RBFN 75-25% partition

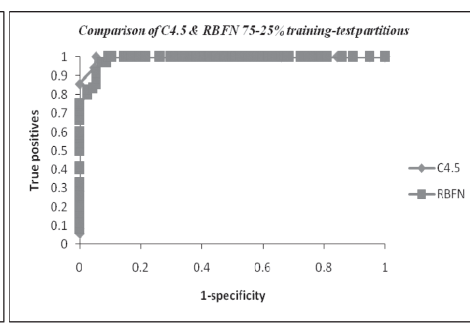

Fig.7. ROC for C4.5 & RBFN 75-25% partition

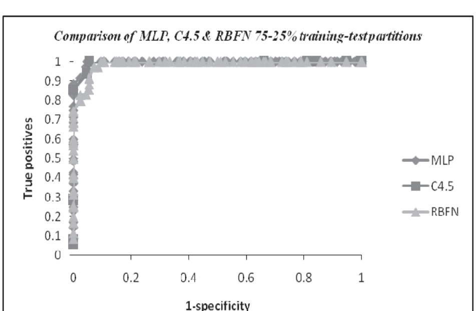

Fig.8. ROC for MLP,C4.5&RBFN 75-25% partition

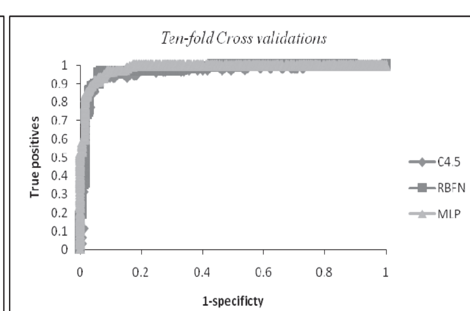

Fig.9. ROC curve for MLP, C4.5& RBFN 10 fold-CV

Figures 2-9 represents the ROC graphs which clearly epitomize the superior performances of the three benchmarked algorithms in this field, where the value generated by MLP-C4.5-RBFN are 0.995-0.957-0.982 respectively using 75-25% training-test partitions and 0.98-0.94-0.97 respectively using ten-fold cross-validations. All the above findings of the current study corroborates a great deal of the earlier work in thyroid classification.

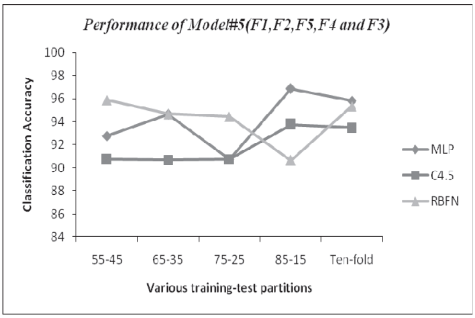

Fig.10. Performance without feature selection

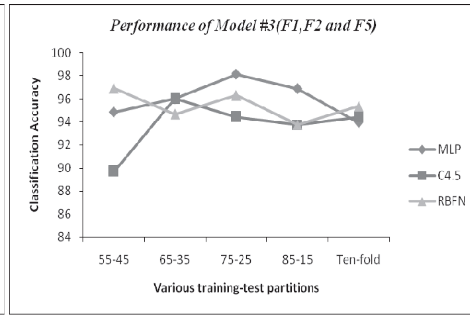

Fig.11. Performance of Model #3 (F1,F2&F5)

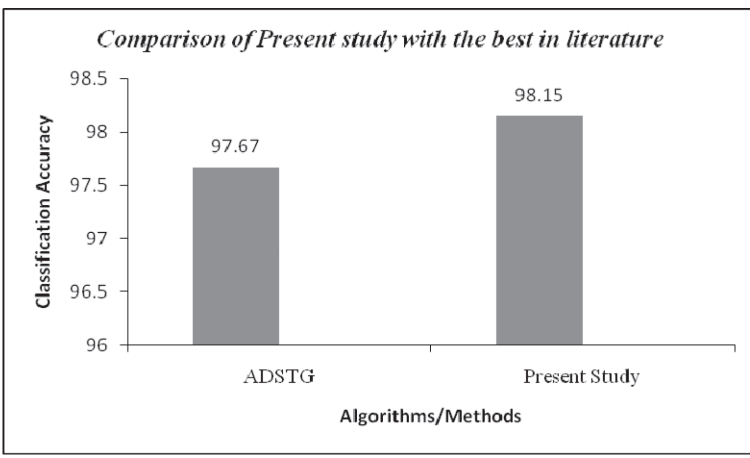

Fig. 12. Comparison of Present study with ADSTG

From the Figure 10 and 11, a comparison of classification accuracies obtained by all the features and selected features can be studied. From the graphs presented in Figure12 it is proved strongly that the current study has outperformed the best result in this field done by ADSTG method.

CONCLUSIONS

This research has investigated thyroid dataset with ranked improved F-score ordering. The study was designed to determine the effect of the proposed ranked improved F-score ordering method for thyroid dataset classification. The result of this investigation has shown that combination of features F1, F2, F5 emerged as reliable predictors of accurate classification of patients. An implication of this is that more number of patients would be diagnosed correctly and early. Also, the most striking is multilayer perceptron (98.15%) produced higher performance together with the proposed method as against C4.5 (94.40), RBFN (96.29%) and ADSTG (97.67%). This research will serve as a base for future studies in classification of medical datasets. The methods used for this study may be applied to other medical datasets and other domains too. Further investigations are needed to estimate the proposed method to explore with large number of patient records and many classes. Taken together, the results obtained suggest that multilayer perceptron is one such classifier, where promising results can be achieved in medical diagnosis. The RIFO has helped to strongly determine the set of combination of features, which produced the best results among results in the literature. The present study, however, made several noteworthy contributions to the field of thyroid dataset exploration and classification, which would be a great help to physicians, patients and the research community.

REFERENCES

Barbaram, B. P., Carrasco-Ochoa, J.A. & Fco, J. 2011. Martinez-Trinidad, General framework for class-specific feature selection, Expert Systems with Applications **38**:10018–10024.

Christopher Jennison. & Bruce W. Turnbull. 2013. Interim monitoring of clinical trials: Decision theory, dynamic programming and optimal stopping. Kuwait J. Sci. **40**(2) pp.43-59.

Coomans, D., Broeckaert, M. Jonckheer M., & Massart D.L. 1983. Comparison of Multivariate Discriminant Techniques for Clinical Data - Application to the Thyroid Functional State. Meth. Inform. Med. **22**:93-101.

Delen, D., Walker. G. & Kadam. A. 2005. Predicting breast cancer survivability: A Comparison of three data mining methods. Artificial Intelligence in Medicine. **34**:113- 127.

Dogantekin, E., Dogantekin, A. & Avci, D. 2010. 'An automatic diagnosis system based on thyroid gland: ADSTG', Expert Systems with Applications, **37**(9) pp.6368–6372.

Dogantekin, E., Dogantekin, A. & Avci, D. 2011. An expert system based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for diagnosis of thyroid diseases, Expert Systems with Applications, **38**:146-150.

Du, K.-L. & Swamy, M.N.S. 2006. Neural Networks in a Soft computing Framework Springer-Verlag.

Feyzullah Temurtas. 2009. A comparative study on thyroid disease diagnosis using neural networks', Expert Systems with Applications, **36**: 944-949.

Fu, X.J. & Wang, L.P. 2003. Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance, IEEE Transactions on Systems, Man and Cybernetics-Part B, **33**: 399-409.

Hoshi, 2005. An analysis of thyroid function diagnosis using Bayesian-type and SOM- type neural networks, Chemical and Pharmaceutical Bulletin, **53**:1570–1574.

Jaganathan, P. & Rajkumar, N. 2012. An expert system for optimizing thyroid disease diagnosis, Int. J. Computational Science and Engineering, **7**: 232-238

Keles, A. & Keles, A. 2008. ESTDD: Expert system for thyroid diseases Diagnosis, Expert Systems with Applications. **34**: 242–246.

Shu-Hsien, L., Chu, P. & Hsiao, P. 2012. Data mining techniques and applications –A decade review from 2000 to 2011, Expert Systems with Applications. **39**:11303–11311

Mitchell,T. M. 2003. Machine learning, China Machine Press, Beijing.

Nahar, J., Imam, T., Tickle, K.S. & Chen, Y.P. 2013. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, Expert Systems with Applications, **40**: 96-104

Niloofar, P., Ganjali, M. & Farid Rohani, M.R.2013. Improving the performance of Bayesian networks in non-ignorable missing data imputation. Kuwait Journal of Science. **40**(2) pp.83-101.

Ozyılmaz, L. & Yıldırım, T. 2002. Diagnosis of thyroid disease using artificial neural Network methods in: Proceedings of ICONIP'02 ninth international conference onneural information processing, Orchid Country Club, Singapore.

Pasi, L. 2004. Similarity classifier applied to medical data sets, 10 sivua, Fuzziness in Finland'04 in: International conference on soft computing, Helsinki, Finland & Gulf of Finland & Tallinn, Estonia.

Paliwal, M. & Kumar, U.A. 2009. Neural networks and statistical techniques: A review of applications, Expert Systems with Applications, **36**:2–17

Polat, K., Sahan, S. & Gunes, S. 2007. A Novel hybrid method based on Artificial Immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid Disease diagnosis, Expert Systems with Applications, **32**: 1141–1147.

Quinlan, J.R. 1996. Improved use of continuous attributes in C4.5, Journal of Artificial Intelligence Research, **4**: 77–90.

Serpen, G., Jiang, H. & Allred, L. 1997. Performance analysis of probabilistic potential function neural network classifier in: Proceedings of artificial neural networks in engineering conference, St.Louis, MO. **7**: 471 – 76.

Xie, J. & Wang, C. 2011. Using support vector machines with a novel feature selection method for diagnosis of erythemato-squamous diseases, Expert Systems with Applications, **38**:5809- 5815.

Yilmaz, I. & Kaynar, O. 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils, Expert Systems with Applications, **38**: 5958–5966.

Zhang, G. & Berardi, L.V. 1998. An investigation of neural networks in thyroid function Diagnosis, Health Care Management Science. **1**:29–37.

Submitted **:** 9/01/2014

Revised **:** 3/02/2014

Accepted **:** 30/03/2014

البناء الأمثل من نماذج تصنيف متعددة لتشخيص مشاكل الغدة الدرقية عند البشر

* ن. راجكومار ، ** ب. جاقاناثان

*أستاذ مساعد – قسم تطبيقات الكمبيوتر – كلية بسنا للهندسة والتكنولوجيا – الهند

** أستاذ – قسم تطبيقات الكمبيوتر – كلية بسنا للهندسة والتكنولوجيا – الهند

خلاصة

مشكلة اضطراب الغدة الدرقية صحية عامة ورئيسية. والكشف المبكر عن اضطراب الغدة الدرقية ذو أهمية متزايدة في مجال التشخيص الطبي والتعرف على الأنماط والتعلم الآلي وتنجيم البيانات. إن اضطراب الغدة الدرقية – سواء في زيادة الإنتاج (فرط الهرمون) أو أقل الإنتاج (قصور الغدة الدرقية) – يسبب حالة من عدم التوازن في تحفيز هرمون الغدة الدرقية عند البشر. لذلك، أصبحت السيطرة على هذا المرض قضية مركزية في مجال الرعاية الصحية وتحتاج إلى عناية كبيرة.

يدرس هذا البحث بشكل نقدي نماذج تصنيف مختلفة شيدت باستخدام ترتيب رياض محسن ومطور للنتيجة -F، والمطبق على قاعدة بيانات الغدة الدرقية التي اتخذت من مستودع التعلم الآلي بجامعة كاليفورنيا في ايرفاين. إجمالي ما تم استخدامه من قاعدة البيانات هو تسع مجموعات فرعية ممكنية وفعالة، يتم اختبار كل مجموعة فرعية بواسطة ثلاثة خوارزميات الأكثر قياسها وهي C4.5 والمستقبلات متعددة الطبقات والشبكة الوظيفية بالدالة ذات القاعدة الدائرية، وقد تم التحقق من صحة النتائج 10 مرات بعد تقسيم البيانات إلى نوعين تدريبية واختبارية. وقد أعطت النتائج التي تم الحصول عليها استنتاجات مختلفة، ولكن واحدة منها عالية الدقة ومثيرة للاهتمام. ومن النتائج لوحظ أيضاً أن المستقبلات متعددة الطبقات أعطت أفضل أداء متميز بنسبة 98.15 %، وهو أعلى من كل البحوث السابقة. قاعدة البيانات لديها 3 فئات، 5 ميزات و215 سجلات (قصور =30، فرط =35 ، عادي =150).