British Journal of Applied Science & Technology
10(1): 1-13, 2015, Article no.BJAST.17613
ISSN: 2231-0843

SCIENCEDOMAIN

SCIENCEDOMAIN ***international***
www.sciencedomain.org

Ensemble Classifier System for Automatic Diagnosis of Thyroid Disease

A. H. El-Baz[1*]

[1]*Department of Mathematics, Faculty of Science, Damietta University, New Damietta, Egypt.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/BJAST/2015/17613
Editor(s):
(1) Qing-Wen Wang, Department of Mathematics, Shanghai University, P.R. China.
(2) A. A. Hanafi-Bojd, Department of Medical Entomology & Vector Control, School of Public Health, Tehran University of Medical Sciences, Iran.
Reviewers:
(1) Anonymous, University of São Paulo, Brazil.
(2) Anonymous, Internal Medicine, Medical University, Zaporozhye, Ukraine.
(3) Anonymous, Giresun University, Turkey.
(4) Rajendra Nath, Pharmacology & Therapeutics, King George's Medical University, Lucknow, India.
Complete Peer review History: http://sciencedomain.org/review-history/9750

Original Research Article

Received 20th March 2015
Accepted 13th May 2015
Published 13th June 2015

ABSTRACT

Thyroid gland secretes hormones that govern many of the functions in our body, such as the way the body uses energy, consumes oxygen and produces heat. Thyroid disorders typically occur when this gland releases too many or too few hormones. An overactive or underactive thyroid can lead to a wide range of health problems. Automatic diagnosis of Thyroid disease via proper interpretation of the thyroid data set is an important classification problem. Thyroid disease dataset which is taken from UCI machine learning database was used. The proposed method uses both Multilayer Perceptron (MLP) and Cascade-Forward Back Propagation Network (CFBN) as base classifiers for the proposed combined classifier systems. The combined classifier is based on varying the parameters related to both the design and training of neural network classifiers. The proposed method achieved accuracy value 96.23% for both combined MLP and combined CFBN classifiers. It has been observed that these results are one of the best results compared with results obtained from related previous studies and reported in the UCI web sites. The experimental results obtained show that the proposed combined classifier can be successfully used for diagnosing thyroid disease.

Keywords: Thyroid disease diagnosis; MLP; CFBN; combined classifier; majority vote.

**Corresponding author: E-mail: ali_elbaz@yahoo.com, elbaz@du.edu.eg;*

1. INTRODUCTION

The thyroid gland is defined as a butterfly-shaped gland in the front of the neck. The thyroid gland lies below the Adam's apple. The function of the thyroid is to regulate numerous metabolic processes throughout the body. When the thyroid is its normal size, one can't feel it [1].

The thyroid secretes several hormones which are called thyroid hormones. The most important hormone is thyroxine which is known as T4. Thyroid hormones act in the human body and they have great influencing in metabolism, growth and development, and body temperature. During infancy and childhood, adequate thyroid hormone is crucial for brain development. Thyroid problems may be caused due to disorders of the pituitary gland and hypothalamus tissues which controlled the thyroid gland [1].

The diagnosis of thyroid disease requires clinical examination, suitable interpretation, and complementary investigation of the thyroid data set. Pattern recognition techniques are used as automatic diagnoses of this type of disease.

Given a pattern recognition problem, the traditional approach is to evaluate a set of different algorithms against a representative validation set and select the best one. However, theoretical and empirical evidence suggests that no single ‘‘best’’ method exists for all classification problems. Also, it is always possible to find the ‘‘best’’ method for a given data set and identify the ‘‘best’’ characteristics of a data set for a given method. This suggests that combining the results of classification methods may result in improved generalization.

Ensembles of neural classifiers have been introduced in [2–4]. The base classifiers are different from each other in one or more aspect such as the training parameters, initial weights, network architecture or training algorithm.

In this paper, we enforce diversity in the ensemble by varying the neural network architectures. These parameters are chosen adaptively to generate a pool of classifiers. The final decisions are combined by using the majority voting rule.

The organization of the rest of the paper is as follows: In Section 2, literature review on thyroid database is given. In Section 3, description of the thyroid dataset is given. The proposed combined classifier is given in Section 4. In Section 5, the experimental evaluations of the proposed ensemble technique are given. Section 6 concludes the paper.

2. LITERATURE REVIEW

There are many diagnosis studies of thyroid disease in literature. So far, in these studies, many techniques, which are Multi-Layer Perceptron with Back-Propagation (MLP with BP), Radial Basis Function (RBF), adaptive Conic Section Function Neural Network (CSFNN) [5], Multi-Layer Perceptron (MLP), Learning Vector Quantization (LVQ), Radial Basis Function (RBF), (Probabilistic Potential Function Neural Network (PPFNN) [6], Linear Discriminant Analysis (LDA), C4.5 with default learning parameters (C4.5-1), C4.5 with parameter c equal to 5 (C4.5-2), C4.5 with parameter c equal to 95 (C4.5-3), MLP, DIMLP with two hidden layers and default learning parameters (DIMLP) [7], expert system for thyroid disease diagnosis with neuro-fuzzy classification (ESTDD with NEFCLASS-J) [8], have been used. Table 3 summarizes and compares the proposed technique with other popular techniques on thyroid dataset.

3. THE THYROID DATASET USED IN THIS STUDY

The thyroid gland (see Fig. 1) is the biggest gland in the neck. Production of too little thyroid hormone causes hypothyroidism or production of too much thyroid hormone causes hyper-thyroidism. 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. 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 [5]. The seriousness of thyroid disorders should not be underestimated as thyroid storm (an episode of severe hyperthyroidism) and myxedema coma (the end stage of untreated hypothyroidism) may lead to death in a significant number of cases [9].

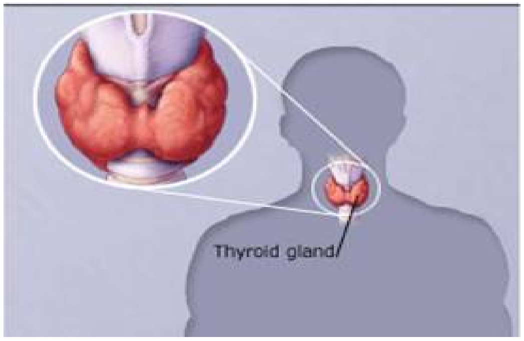

Fig. 1. The Thyroid gland [11]

The thyroid data employed in this study is due to Coomans [10] and is available from UCI Repository site; UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/), which comprises 215 patients from the same hospital. These individuals were divided into three groups of known classification, based on diagnosis results, healthy individual who we call as "normal" in the following for which there were 150 cases, patients suffering from hyperthyroidism (hyper) for which there were 35 cases, from hypothyroidism (hypo) for which there were 30 cases. Table 1 describes the class distribution. The thyroid data are measurement of the thyroid gland. Each individual was characterized by the result of five laboratory tests. Table 1 describes Class distribution of the thyroid dataset.

These are as follows:

- T3-resin uptake test (A percentage).
- Total Serum thyroxin as measured by the isotopic displacement method.
- Total serum tri-iodothyronine as measured by radioimmunoassay.
- Basal thyroid-stimulating hormone (TSH) as measured by radioimmunoassay.
- Maximal absolute difference of TSH value after injection of 200 lg of thyrotropin-releasing hormone as compared to the basal value.

4. METHOD

4.1 Multilayer Perceptron Neural Network

The multilayer neural networks (MLNNs) have been successfully used in replacing conventional pattern recognition methods for the thyroid diagnosis systems [6,7].

In the first stage of this study, a multilayer neural network structure, see Fig. 2, was used for the thyroid disease diagnosis. The network structure consists of (input layer, hidden layers, and output layer). The hidden layer neurons and the output layer neurons use nonlinear sigmoid activation functions. In this system, five inputs were features, and three outputs are index of three classes.

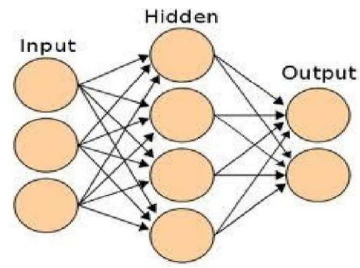

Fig. 2. Feed forward back propagation network

Levenberg–Marquardt algorithm [12] was used in this study for the training of the multilayer neural network structure.

4.2 Cascade-forward Back Propagation Network

A cascade type of feed-forward ANNs, see Fig. 3, consists of three layers as follow: input layer, output layer, and one or more hidden layers to enable it to solve nonlinear problems. The first layer has weights receiving from the input. But each subsequent layer has weights coming from the input and all previous layers. All the network layers have biases. The last layer is for the output of the network. All weights and biases must be initialized.

Generally, the trainable cascade-forward back-propagation network models are similar to feed-forward networks, but include a weight connection from the input to each layer and from each layer to the successive layers. Filik and Kurban [13] were found that cascade-forward back-propagation method can be more efficient than feed-forward back-propagation method in the some cases.

Table 1. Classes distribution of the thyroid dataset

Index	Class name	Class size	Class distribution (%)
C1	Normal	150	69.77
C2	Hyper	35	16.28
C3	Hypo	30	13.95

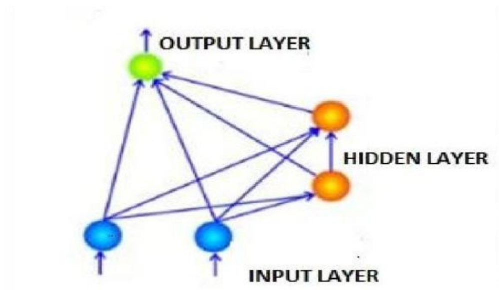

Fig. 3. Cascade forward back propagation network

4.3 Proposed Combined Classifier System

Combining multiple classifiers has been intensively studied and widely regarded as an effective technique for overcoming the limitations of individual classifiers [14,15]. Importantly, these classifier ensemble methods, e.g., neural network ensembles, have been applied in many real-world applications and achieved great success [16,17].

Generally speaking, there are two main categories in the classifier ensemble. The first one aims at learning multiple classifiers at the feature level. Namely, they usually train and combine multiple classifiers in the learning process, e.g., Boosting [18] and Bagging [19]. On the other hand, the second school of methods tries to combine classifiers at the output level, where the results of multiple available classifiers are combined to solve the targeted problem, e.g., multiple classifier systems, or mixture of experts [20]. In this paper, we focus on the second one. Namely, given multiple classifiers, the classifier ensemble is learned by combining these component classifiers intelligently. Furthermore, classifiers differing in feature representation, architecture, learning algorithm, or training data exhibit complementary classification behavior and the fusion of their decisions can yield higher performance than the best individual classifier.

Also, it had been observed in design studies that although one of the designs would yield the best performance, the sets of patterns misclassified by different classifiers would not necessarily overlap. This suggested that different classifier designs potentially offered complementary information about the patterns to be classified, which could be harnessed to improve the performance of the selected classifier.

On the other hand, to build a multiple classifier system, there are many approaches. In [21], three of these approaches have been presented: divide and conquer approach, sequential approach and parallel approach. According to [21], in the divide and conquer approach the types of input are isolated on which each specific classifier performs well, and new input is directed accordingly. While in the sequential approach, one classifier is used at the beginning and other classifiers are invoked only if the first classifier is failed to satisfy the proposed result. In the parallel approach, which is applicable in many studies and experiments, all individual classifiers are applied in parallel to the same input data, and a combination method is applied to combine their decisions. This technique has been used in this paper.

Mainly, we train randomly many individual classifiers. Each classifier is different than the others in the number of epochs, the learning rate, and/or the number of hidden neurons. Then we get the hypotheses of these classifiers on the dataset. At the end, we combine the classifiers results by using the majority voting technique (also known as plurality voting technique). Both multilayer neural network and cascade-forward back propagation network are used as base classifier in the proposed ensemble systems. The topology of the parallel combination method and the flowchart given in Figs. 4 and 5, respectively show briefly the proposed ensemble method of the individual classifiers.

4.3.1 Majority vote combiner

In this study, a majority voting, which is a popular way of combining classifiers, is used. Majority counts the votes for each class over the input classifiers and select the majority class [22,23]. It is operation algorithm is as follow [24]:

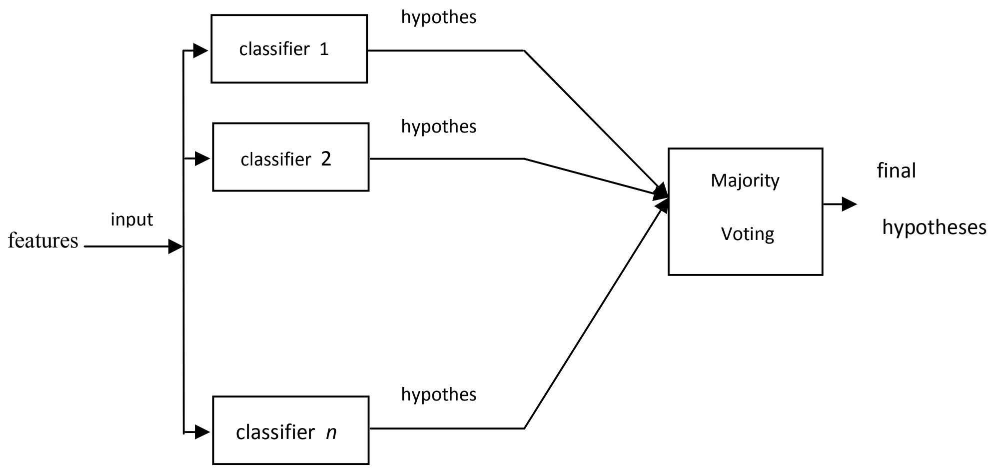

Fig. 4. The topology of the parallel combination of the proposed method

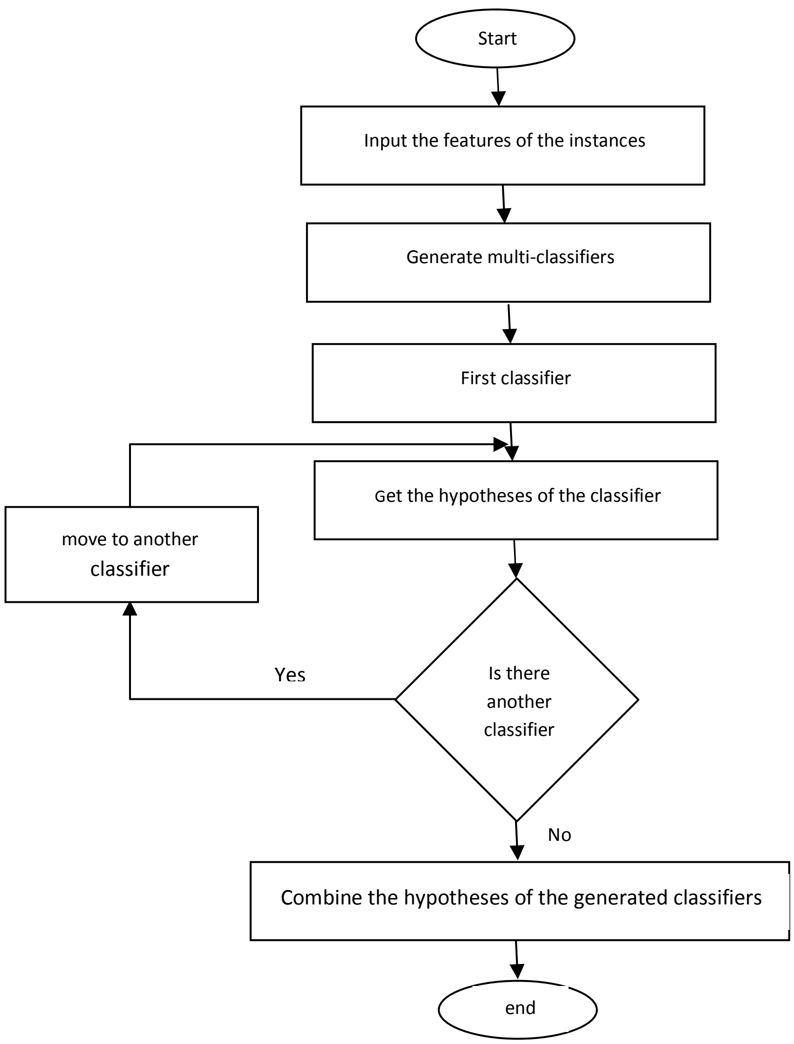

Fig. 5. Flowchart for the proposed ensemble method

For each new pattern, find the class labels $y_1,...,y_n$, assigned to this pattern by the *n* base classifiers. Next, calculate the number of votes for each class d_k, $k=1,...,c$. $N(k)=\sum_{i=1}^{n} I(y_i,d_k)$, where $I(\alpha,\beta)=1$ if $a=b$ and 0 otherwise. After that, assign label k^* to the object, where $k^*=\arg\max_{k=1}^{c} N(K)$. Finally, return the ensemble label of the new object.

5. EXPERIMENTAL RESULTS

Table 2 summarizes the number of training and testing data for each classes of the thyroid dataset used in our experiment.

5.1 Results of Applied Ensemble of Multilayer Perceptron (MLP) on Thyroid Database

In the first experiment, MLP is used as base classifier for the proposed ensemble. For the sake of generate diversity in classifier ensemble, a pool of 16 different classifiers is generated and the final hypothesis is combined using majority voting technique. The percentage of correct classification (PCC) of the best individual classifier on the test data is 94.3396% and the PCC of the proposed combined classifier on the test data is 96.2264%. We note that the proposed combined classifier compared favorably with the results stated in literatures, see Table 3 and Figs. 14-16. Fig. 6 shows the accuracy of base MLP versus the number of individual classifiers. Also, Figs. 7,8 and 9 show the Performance of the MLP neural network, Training state of MLP and Regression of MLP, respectively.

Table 2. The number of training and test data for each class

	The number of training data (75%)	The number of test data (25%)
The hyper-function class (class-1)	113	37
The hypo-function class (class-2)	26	9
The normal-function class (class-3)	23	7
Total	162	53

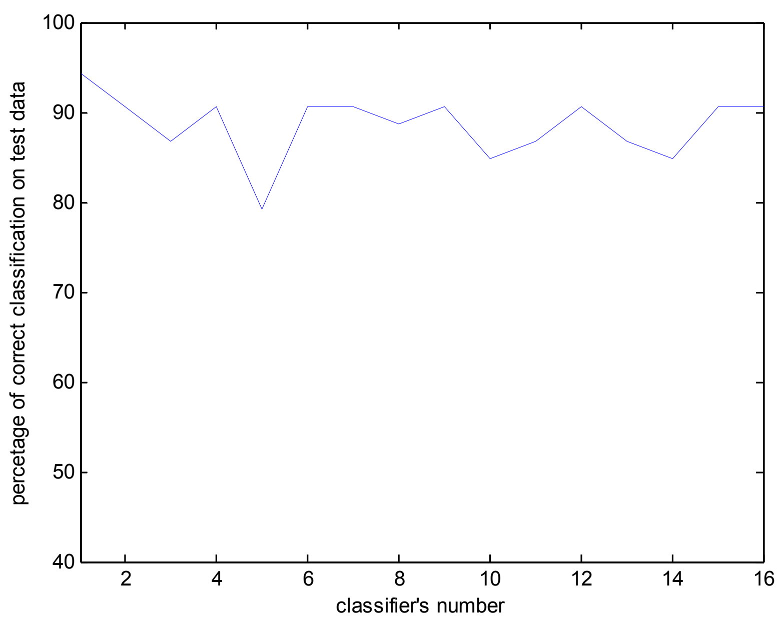

Fig. 6. Accuracy of base MLP versus the number of individual classifier

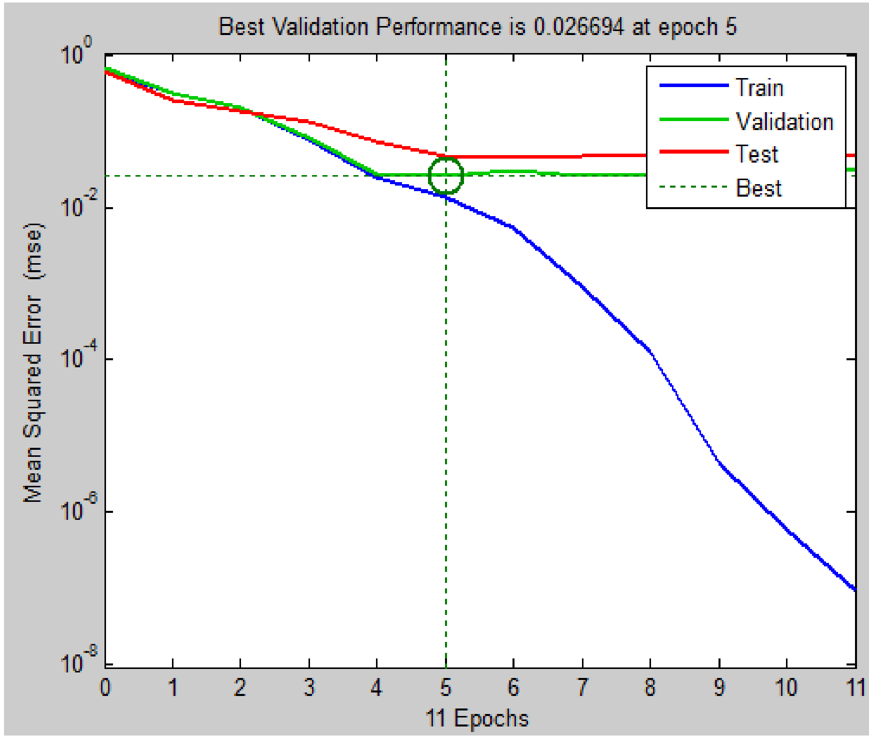

Fig. 7. Performance of the MLP neural network

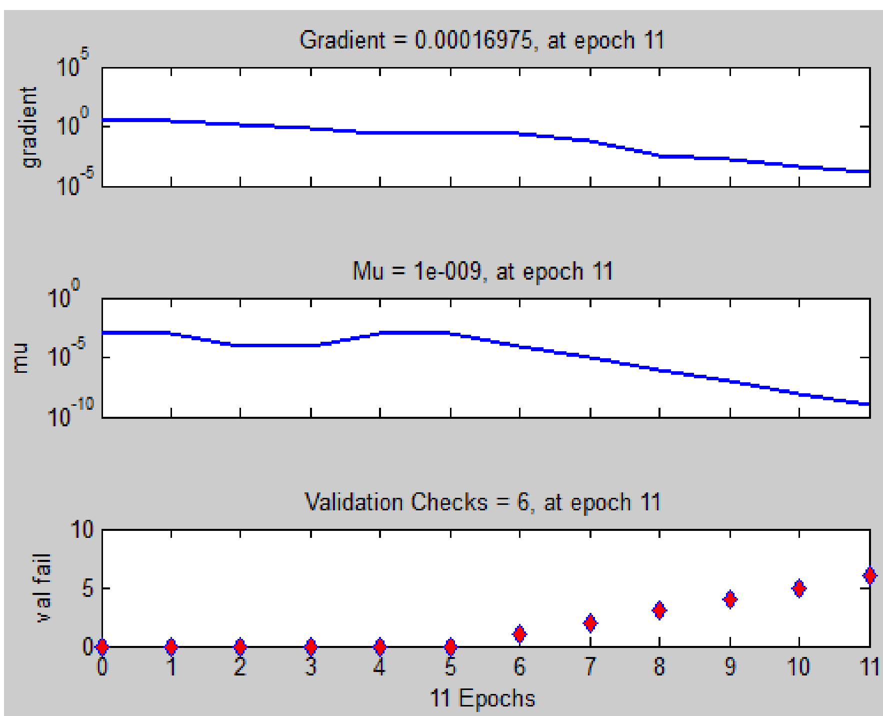

Fig. 8. Training state of MLP neural network

5.2 Results of Applied Ensemble of a Cascade-forward Back Propagation Network on Thyroid Database

In this experiment a cascade-forward back propagation network (CFBN) is used as the base classifier for the combined classifier. Percentage of correct classification (PCC) of best individual classifier, in this case, on the test data is 90.5660% and the PCC of the proposed combined classifier on the test data is 96.2264%. Also, We note that the proposed combined classifier compared favorably with the results stated in literatures see Table 3 and Figs. 14-16. Fig. 10 shows the accuracy of base CFBN versus the number of individual classifier. Also, Figs. 11,12 and 13 show the Performance of the CFBN neural network, Training state of CFBN and Regression of CFBN, respectively.

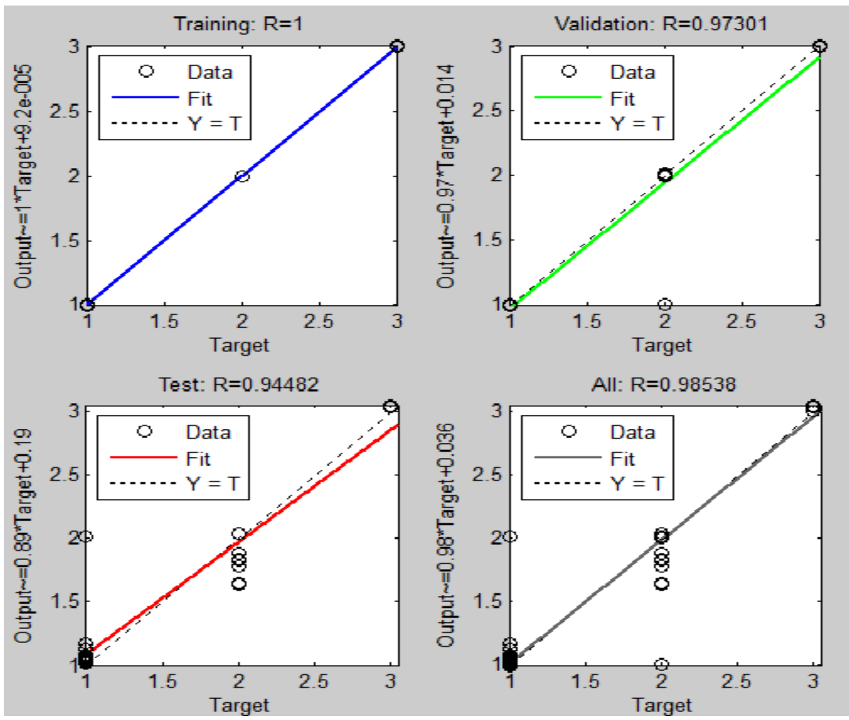

Fig. 9. Regression of MLP neural network

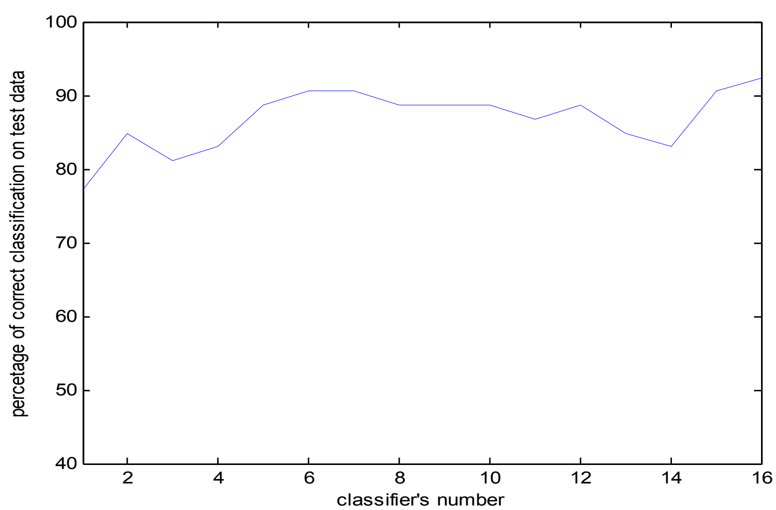

Fig. 10. Accuracy of base CFBN versus the number of individual classifier

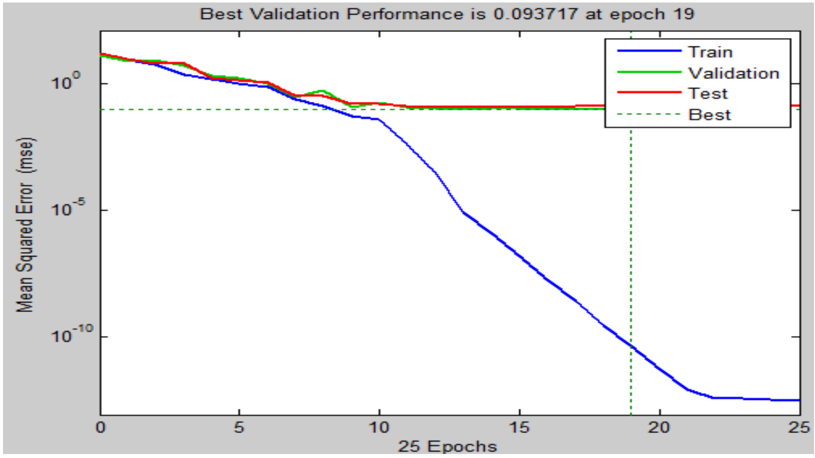

Fig. 11. Performance of the CFBN neural network

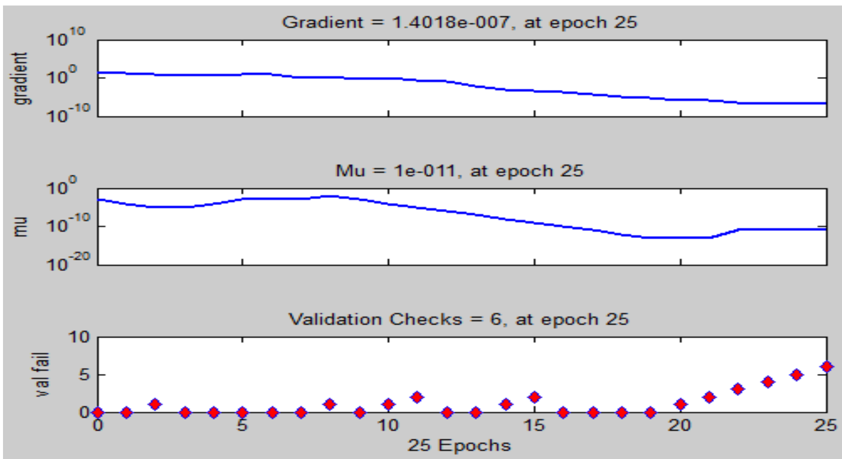

Fig. 12. Training state of CFBN neural network

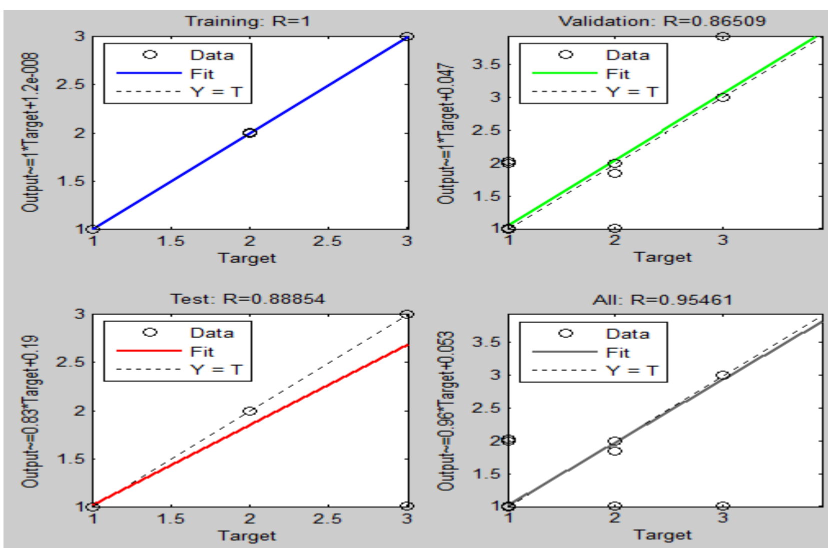

Fig. 13. Regression of CFBN neural network

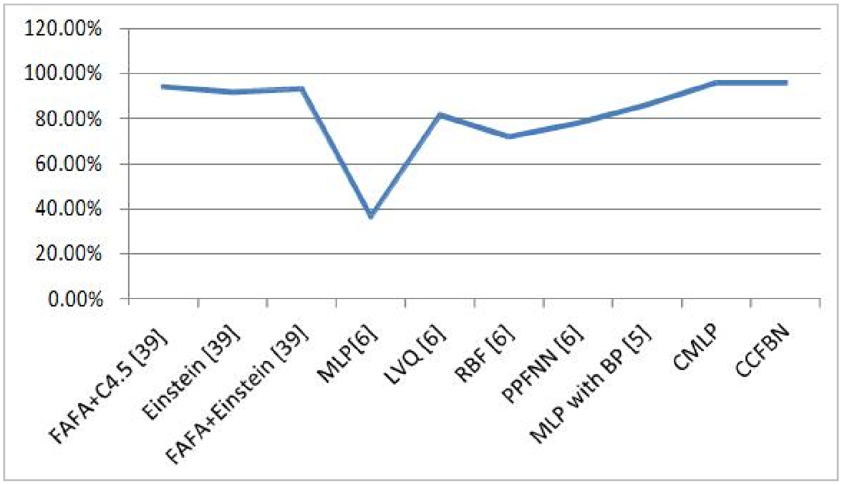

Fig. 14. Comparing the proposed approaches (CMLP and CCFBN) with other approaches

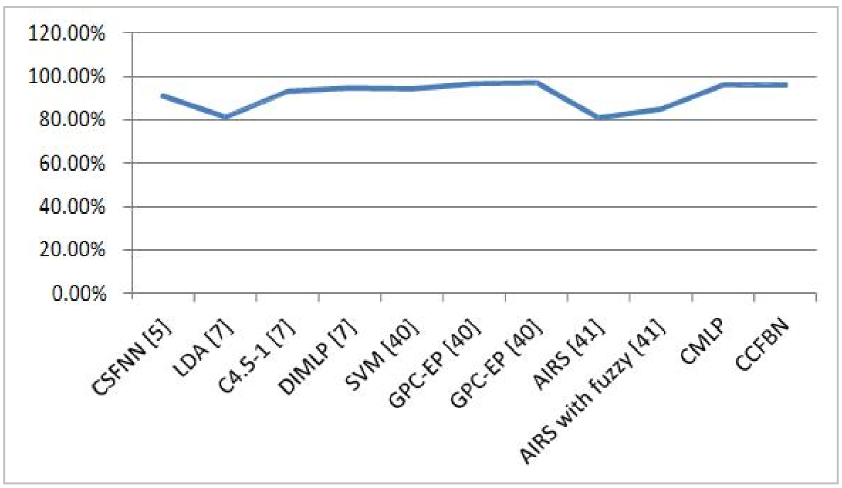

Fig. 15. Comparing the proposed approaches (CMLP and CCFBN) with other approaches

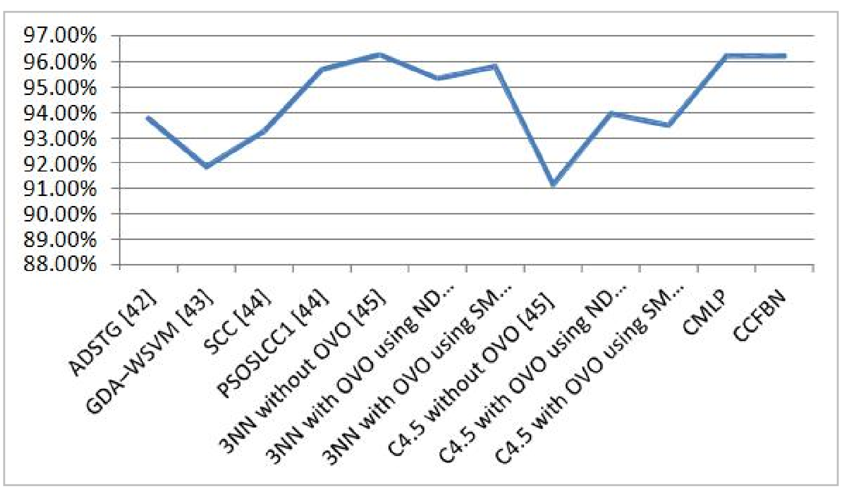

Fig. 16. Comparing the proposed approaches (CMLP and CCFBN) with other approaches

Table 3. Comparison of the obtained results with other results stated in literature for the diagnosis of thyroid disease

Method	Author	Accuracy (%)	References	Year
FAFA+C4.5 (Pruned)	Yip and Webb	94.38 (10-fold cross validation)	[25]	1994
FAFA+C4.5 (Rules)	Yip and Webb	94.38 (10-fold cross validation)	[25]	1994
Einstein	Yip and Webb	91.91 (10-fold cross validation)	[25]	1994
FAFA+Einstein	Yip and Webb	93.34 (10-fold cross validation)	[25]	1994
Multi-Layer Perceptron (MLP)	Serpen et al.	36.74 (test data)	[6]	1997
Learning Vector Quantizer (LVQ)	Serpen et al.	81.86 (test data)	[6]	1997
Radial Basis Function (RBF)	Serpen et al.	72.09 (test data)	[6]	1997
Probabilistic Potential Function Neural Network (PPFNN)	Serpen et al.	78.14 (test data)	[6]	1997
Multi-Layer Perceptron with Back-Propagation (MLP with BP)	Özyılmaz and Yıldırım	86.33 (average 3-fold cross validation)	[5]	2002
Radial Basis Function (RBF)	Özyılmaz and Yıldırım	79.08	[5]	2002
Adaptive Conic Section Function Neural Network (CSFNN)	Özyılmaz and Yıldırım	91.138	[5]	2002
Linear Discriminant Analysis (LDA)	Pasi	81.34 (test data)	[7]	2004
C4.5 with default learning parameters (C4.5-1)	Pasi	93.26 (test data)	[7]	2004
C4.5 with parameter c equal to 5 (C4.5-2)	Pasi	92.81 (test data)	[7]	2004
C4.5 with parameter c equal to 95 (C4.5-3)	Pasi	92.94 (test data)	[7]	2004
Multi-Layer Perceptron (MLP)	Pasi	96.24 (test data)	[7]	2004
DIMLP with two hidden layers and default learning parameters (DIMLP)	Pasi	94.86 (test data)	[7]	2004
LDA	Kim and Ghahramani	93.44 (10-fold cross-validation)	[26]	2006
SVM (soft)	Kim and Ghahramani	94.44 (10-fold cross-validation)	[26]	2006
GPC-EP (s,soft)	Kim and Ghahramani	96.75 (10-fold cross-validation)	[26]	2006
GPC-EP (m,soft)	Kim and Ghahramani	97.23 (10-fold cross-validation)	[26]	2006
The artificial immune recognition system (AIRS)	Polat et al.	81.00 (10-fold cross-validation)	[27]	2007
The artificial immune recognition system (AIRS) with fuzzy weighted pre-processing	Polat et al.	85.00 (10-fold cross-validation)	[27]	2007
ADSTG method	Dogantekin et al.	93.77 (10-fold cross-validation)	[28]	2010
GDA–WSVM expert system	Dogantekin et al.	91.86	[29]	2011
SCC	Liu et al.	93.27	[30]	2014
PSOSLCC1	Liu et al.	95.70	[30]	2014
3NN as base classifier without OVO strategy	Galar et al.	96.28	[31]	2014
3NN with OVO strategy using ND aggregations	Galar et al.	95.35	[31]	2014
3NN with OVO strategy using SM aggregations	Galar et al.	95.81	[31]	2014
C4.5 as base classifier without OVO strategy	Galar et al.	91.16	[31]	2014
C4.5 with OVO strategy using ND aggregations	Galar et al.	93.95	[31]	2014
C4.5 with OVO strategy using SM aggregations	Galar et al.	93.49	[31]	2014
Proposed method (CMLP)		96.23 (test data)		
Proposed method (CCFBN)		96.23 (test data)		

6. CONCLUSION

In this study, combined classifier method (CCM) was presented for diagnosis of thyroid disease. Then, the performance evaluation of CCM for diagnosis of thyroid disease was estimated by using classification accuracy. It is concluded that the CCM for diagnosis of thyroid diseases obtains very promising results in classifying the possible thyroid patients. This status of CCM for diagnosis of thyroid diseases is clearly seen from the obtained results. It is believed that the proposed CCM diagnosis method used in this study can be very helpful to the physicians for their final decision on their patients. The physicians can make very accurate decisions by using such an efficient tool. In future studies of thyroid diagnostic, different feature selection and ensemble classifier methods will be used for the possibility of increasing the accuracy.

ACKNOWLEDGEMENTS

The author thanks anonymous reviewers for providing insightful comments which improve the quality of the manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Available:http://www.medicinenet.com/thyroid_disorders/article.htm
2. Giacinto G, Roli F. Design of effective neural network ensembles for image classification. J Image Vis Comp. 2001; 19(9–10):697–705.
3. Sharkey AJC. On combining artificial neural nets. Connection Sci. 1996;8:299–314.
4. Sharkey AJC, Sharkey NE, Gerecke U, Chandroth GO. The test and select approach to ensemble combination. In: Kittler J, Roli F (eds) Proceedings of the first international workshop on multiple classifier systems (MCS2000) LNCS 1857. Springer. 2000;30–44.
5. Özyılmaz L, Yıldırım T. Diagnosis of thyroid disease using artificial neural network methods. In Proceedings of ICONIP’02 ninth international conference on neural information processing, Orchid Country Club, Singapore. 2002;2033–2036.
6. Serpen G, Jiang H, Allred L. Performance analysis of probabilistic potential function neural network classifier. In Proceedings of artificial neural networks in engineering conference, St. Louis MO. 1997;7:471–476.
7. Pasi L. 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; 2004.
8. Keles A, Keles A. ESTDD: Expert system for thyroid diseases diagnosis. Expert Systems with Applications. 2008;34(1): 242–246.
9. Zhang G, Berardi LV. An investigation of neural networks in thyroid function diagnosis. Health Care Management Science. 1998;29–37.
10. Coomans D, Broeckaert I, Jonckheer M, Massart DL. Comparison of multivariate discrimination techniques for clinical data—Application to the thyroid functional state. Methods Information in Medicine. 1983;22:93-101.
11. Available:http://www.webmd.com/women/guide/understanding-thyroid-problems-basics
12. Hagan MT, Menhaj M. Training feed forward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks. 1994;5:989–993.
13. Filik UB, Kurban M. A new approach for the short-term load forecasting with autoregressive and Artificial Neural Network Models. International Journal of Computational Intelligence Research. 2007;3(1):66–71.
14. Ho T, Hull J, Srihari S. Decision combination in multiple classifier systems. IEEE Trans. Pattern Anal. Mach. Intell. 1994;16(1):66–75.
15. Dietterich T. Ensemble method in machine learning, in: Proceedings of International Workshop on Multiple Classifier Systems. 2000;1–15.
16. Dasarathy BV. A special issue on applications of ensemble methods. Inf. Fusion. 2008;9(1):1-140.
17. Oza C, Tumer K. Applications of ensemble methods. Inf. Fusion. 2008;9(1):2–3.
18. Schapire R. The strength of weak learnability. Machine Learn. 1990;5(2): 197–227.

19. Breiman L. Bagging predictors. Machine Lear. 1997;24(2):123-140.
20. Yuksel SE, Wilson JN, Gader PD. Twenty years of mixture of experts. IEEE Trans. Neural Network Learn. Syst. 2012;23(8): 1177-1193.
21. Ho TK. Data complexity analysis for classifier combination, in: Proceedings of the 2nd International Workshop of Multiple Classifier System, Cambridge, UK, Lecture Notes in Computer Science, Springer-Verlag, 2096. 2001;53-67.
22. Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998;20(3):226-239.
23. Xu L, Krzyzak A, Suen CY. Methods of combining multiple classifiers and their application to handwriting recognition. IEEE Trans. SMC. 1992;22:418-435.
24. Kuncheva L. Combining pattern classifiers: Methods and algorithms. 2nd ed., John Wiley; 2014.
25. Yip SP, Webb GI. Empirical function attribute construction in classification learning. Joint conference on artificial intelligence (AI.'94). 1994;29-36.
26. Kim HC, Ghahramani Z. Bayesian Gaussian process classification with the EM-EP algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006;28(12):1948–1959.
27. Polat K, Sahan S, Gunes S. A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis. Expert Systems with Applications. 2007;32:1141–1147.
28. Dogantekin E, Dogantekin A, Avci D. An automatic diagnosis system based on thyroid gland: ADSTG. Expert Systems with Applications. 2010;37:6368-6372.
29. Dogantekin E, Dogantekin A, Avci D. An expert system based on generalized discriminant analysis and wavelet support vector machine for diagnosis of thyroid diseases. Expert Systems with Applications. 2011;38:146-150.
30. Liu R, Chen Y, Jiao L, Li Y. A particles warm optimization based simultaneous learning framework for clustering and classification. Pattern Recognition. 2014; 47:2143-2152.
31. Galar M, Fernandez A, Barrenechea E, Herrera F. Empowering difficult classes with a similarity-based aggregation in multi-class classification problems. Information Sciences. 2014;264:135-157.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/9750