ANADOLU ÜNİVERSİTESİ

Bilim ve Teknoloji Dergisi A-Uygulamlı Bilimler ve Mühendislik
Cilt: 15 Sayı: 1 2014
Sayfa: 41-49

ARAŞTIRMA MAKALESİ / **RESEARCH ARTICLE**

Yılmaz KAYA [1]

A FAST INTELLIGENT DIAGNOSIS SYSTEM FOR THYROID DISEASES BASED ON EXTREME LEARNING MACHINE

ABSTRACT

With iodine taken from outside, the thyroid gland is an organ that secretes hormones called thyroxin. All metabolic functions of human beings are controlled by these hormones. An overactive thyroid gland which is producing an excessive amount of these hormones causes hyperthyroidism, while an underactive thyroid gland that is not producing enough of these hormones causes hypothyroidism. The diagnosis of thyroid gland disorders by assessing the data of thyroid in clinical applications comes out as an important classification problem. In this study, Extreme Learning Machine (ELM) was applied to the thyroid data set taken from UCI machine learning repository. The ELM is single hidden layer feed-forward artificial neural network model which can be learnt fast. It was seen that the ELM, for the data set, has the upper hand in terms of both classification accuracy and speed when compared to other machine learning methods. The classification accuracy obtained through the ELM is 96.79% for 70-30% training-test partition.

Keywords: Extreme Learning Machine, Thyroid Diseases, Machine Learning, Expert System.

TİROİT HASTALIKLARIN TEŞHİSİ İÇİN AŞIRI ÖĞRENME MAKİNESİ TABANLI UZMAN BİR TANI SİSTEMİ

ÖZ

Tiroit bezi dışarıdan alınan iyot minerali ile "tiroksin" denilen hormonları yapan bir organdır. İnsana ait tüm metabolizma faaliyetleri bu hormonları tarafından kontrol edilmektedir. Bu hormonların aşırı salınması hyperthyroidism, az salınması ise hypothyroidism bozuklarının ortaya çıkmasına neden olmaktadır. Klinik uygulamalarda tiroit verilerin yorumlanarak tiroit bezi bozukluğu tanısının konulması önemli bir sınıflandırma problemi olarak karşımıza çıkmakta. Bu çalışmada UCI makine öğrenmesi veri tabanından alınan tiroit veri setine aşırı öğrenme makinesi (AÖM) yöntemi uygulanmıştır. AÖM hızlı öğrenebilen tek gizli katmanlı ileri beslemeli bir yapay sinir ağ modelidir. Ele alınan veri seti için AÖM, diğer makine öğrenmesi yöntemlere göre hem sınıflandırma başarısı hem de hız bakımından önemli avantajlar sağladığı görülmüştür. AÖM ile elde edilen sınıflandırma başarısı 96.79 % olarak elde edilmiştir.

Anahtar Kelimeler: Aşırı Öğrenme Makinesi, Tiroit Hastalıklar, Makine Öğrenmesi, Uzman Sistem.

[1,] Siirt University, Engineering and Architecture Faculty, Computer Engineering Department, 56100 Siirt, Turkey.
E-mail: yilmazkaya1977@gmail.com, Tel: +90 506 488 49 90

Received:15 October 2012 **Revised:**30 November 2013 **Revised:**11 December 2014 **Accepted:**15 May 2014

1. INTRODUCTION

The thyroid gland is located in the fore neck just below the bulge called the Adam's apple, and it is an endocrine organ that regulates body metabolism with the hormones it secretes (Özyılmaz and Yıldırım, 2002). From the frontal view it seems like a butterfly as well as "U" shape. After iodine taken through foods passes into blood, it reaches to the thyroid gland, and is used in secreting T3 and T4 hormones (Polat et al., 2007). All cells in our body are affected by T3 and T4 hormones being carried through blood. All metabolic functions of a human are controlled with thyroid hormones. That organ is needed throughout the whole life, and any disorder with it must not be underestimated. The most common diseases with thyroid gland are goitre, thyroid cancer, hyperthyroidism, hypothyroidism, thyroiditis, and solitary thyroid nodules (Esin et al., 2011).These diseases can be explained in sum as below:

Hyperthyroidism, also called thyrotoxicosis, is a disease caused by overproduction of thyroid hormones (T4 & T3). One of the most important causes of Hyperthyroidism is Graves' disease. Graves' disease is a disease of unknown cause of thyroid gland. In the disease the body produces TSH-receptor antibodies against thyroid gland, and these antibodies cause an excessive amount of thyroid hormones producing by stimulating thyroid (Özyılmaz and Yıldırım, 2002).

Hypothyroidism is a condition in which the thyroid gland does not produce enough T3 and T4 hormones. Hyperthyroidism and Hypothyroidism are common thyroid disorders.

Goitre, mostly, does not have any symptom apart from a swelling in the thyroid gland. Because of the enlargement, there could be compression on the local structures; and breathing or swallowing difficulties can be seen.

Thyroid nodule is a serious disease and it necessitates energetic medical intervention if it is seen particularly in men. The thyroid nodule is not different from other organ masses. Each organ mass should be taken serious and should be examined meticulously to learn whether it is a cancer mass or not; and then should be treated. The incidence of thyroid cancer is common cancer type among people. So, the correct thyroid disorder diagnosis by assessing clinical thyroid data is crucial.

Today in the diagnosis of thyroid disorders, machine learning methods as well as clinical studies are used widespread. Artificial neural networks (Zhang and Berardi, 1998; Hoshi et al., 2005; Temurtas, 2009), artificial immune systems (Halife et al., 2009), neuro fuzzy models (Keles and Keles, 2008), decision trees (Pasi, 2004), mixture models (Shu-Kay et al., 2007), stochastic models (Chih-Lin et al., 2010) are common machine learning methods.

In this study, data set taken from UCI repository was used. Extreme Learning Machine (ELM) was used for the classification of the data set. ELM is single hidden layer feed-forward artificial neural network model. With the ELM, the weightings belonging to neurons at the input layer, and the bias values belonging to neurons in the hidden and input layers, are all randomly generated. By contrast, the outputs from the hidden layer are computed analytically (Huang et al., 2006; Kaya, 2013). The most significant feature of the ELM model is that the learning process is very efficient. It can learn thousands of times faster than conventional learning algorithms for feed-forward neural networks. The learning speed of other feed-forward neural networks is typically relatively slow, largely due to the slow gradient-based learning algorithms used in the training procedure (Kaya, 2013). Time, particularly, is a vital factor in diagnosis in medical applications. Artificial neural network (ANN), Naive Bayes (NB), support vector machines (SVM), and decision trees (DT) which are realised on the same data set with ELM got better results in terms of speed and classification performances.

The content of this study was designed as follows. Obtaining and introduction of the dataset were explained in next section. Theoretical information of ELM method was provided in section three. Obtained experimental results were shared in section four. This study was discussed in final section.

2. MATERIAL and METHOD
2.1. Material

Data set was taken from UCI machine learning repository (ftp://ftp.ics.uci.edu/pub/machine-learningdatabases (UCI, 2013). The reason of why this data set was used is its widespread use in classification. Data set compose of total 215 examples, of which 150 normal, 35 hyperthyroidism and 30 hypothyroidism. Obtained clinical features are total 5, they are:

Feature 1: T3-resin uptake test. (A percentage)

Feature 2: Total Serum thyroxin as measured by the isotopic displacement method.

Feature 3: Total serum triiodothyronine as measured by radioimmuno assay.

Feature 4: Basal thyroid-stimulating hormone (TSH) as measured by radioimmuno assay.

Feature 5: Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropin-releasing hormone as compared to the basal value.

2.2. Method

In this study, Extreme Learning Machine (ELM), which is a machine learning method developed by Huang (2006), was used for diagnosis of thyroid diseases. ELM is a single hidden layer feed-forward artificial neural network model which the input weights randomly and output weights are analytically calculated (Yuan et al, 2011; Kaya and Murat, 2013).

Weights in the conventional feed-forward neural networks need to be updated by the gradient based learning algorithms. Yet, to provide a better performance, the procedure of training takes both much time and the error can fall into local optima. Changing the momentum value, maybe, can decrease the risk of the error being trapped in local optima, but it cannot affect the time spent on the training process which lasts long (Suresh et al., 2010; Kaya et al., 2014). Input-output weights and bias values in a single hidden layer feed-forward neural network (SLFN) do have an impact on the performance of the network (Huang et al., 2006). Input-output weights and bias values in ELM are selected randomly; however the output weights are calculated analytically. Therefore, the ELM has shown faster learning speed and better generalization performance on most real tasks than conventional learning algorithms (Kaya, 2013).

2.2.1. Extreme Learning Machine (ELM) Algorithm

The SLFN structure is illustrated in Figure 1.

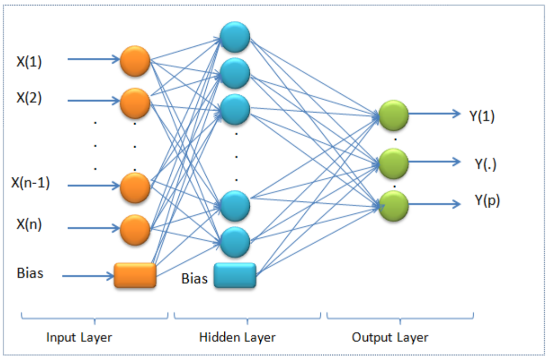

Figure 1. The structure of a single hidden layer feed-forward artificial neural network.

According to Figure 1, on determining that $X = (X_1, X_2, X_3 ..... X_N)$ is input and $Y = (Y_1, Y_2, Y_3 ..... Y_N)$ is output, the mathematical model with M hidden neurons can be defined as in Kaya (2013):

$$\sum_{i=1}^{M} \beta_i g(W_i X_k + b_i) = O_k \quad , \quad k = 1,2,3.......N \quad (1)$$

Where $W_i = (W_{i1}, W_{i2}, W_{i3} ..... W_{in})$ and $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3} ..... \beta_{im})$ are the input and output weights; b_i is the threshold of the hidden neuron and O_k is the output of the network. $g(.)$ denotes the activation function (Hai-Jun; 2006).

In a network of the N training samples, the aim is with zero error: $\sum_{k=1}^{N} (o_k - Y_k) = 0$ or with min error: $\sum_{k=1}^{N} (o_k - Y_k)^2$. Therefore, equation 1 can be shown as below (see Huang ,2006; Kaya et al., 2014).

$$\sum_{i=1}^{M} \beta_i g(W_i X_k + b_i) = Y_k \quad , \; k = 1,2,3.......N \tag{2}$$

Because in the equation above $g(W_i X_k + b_i)$ denotes output matrix in the hidden layer, equation 2 can be placed as (Huang et al., 2006):

$$H\beta = Y \tag{3}$$

This is where;

$$H(W_1,........W_M\,;\, b_1,........b_M\,;\, X_1,.......X_N) = \begin{bmatrix} g(W_1 X_1 + b_1)..... & g(W_M X_M + b_M \\ . & . \\ . & . \\ g(W_1 X_N + b_1)...... & g(W_M X_N + b_M) \end{bmatrix} \tag{4}$$

And

$$\beta = \begin{bmatrix} \beta_1^T \\ . \\ . \\ \beta_M^T \end{bmatrix}_{Mxm} \quad and \quad Y = \begin{bmatrix} Y_1^T \\ . \\ . \\ Y_N^T \end{bmatrix}_{Nxm} \tag{5}$$

H denotes the hidden layer output matrix (Suresh, 2010).

Input weights in ELM $W_i = (W_{i1}, W_{i2}, W_{i3}.....W_{in})$ and b_i hidden layer biases have been randomly produced, and the hidden layer output matrix *H* is obtained analytically. The procedure of training an SLFN is to seek a least-squares solution of the linear system $H\beta = Y$ in ELM (Yuan et al., 2011):

$$\left\| H(W_1, W_2, ...W_M\,; b_1, b_2, ...b_M)\hat{\beta} - Y \right\| = \min_{\beta} \left\| H(W_1, W_2, ...W_M\,; b_1, b_2, ...b_M)\beta - Y \right\| \tag{6}$$

Above, $\hat{\beta} = H^{+}Y$ is the smallest norm least-squares of $H\beta = Y$ Furthermore, H^{+} indicates the Moore-Penrose generalized inverse of *H*. The norm of $\hat{\beta}$ is the smallest among all the least-squares solutions of [14]. ELM algorithm can be summarized in 3 steps as below (Handako et al., 2006; Liang et al., 2006; Zong and Huang, 2006; Paavo and Pasi, 2008).

1. **Phase**: $W_i = (W_{i1}, W_{i2}, W_{i3}.....W_{in})$ input weights and hidden layer b_i bias values are produced randomly.
2. **Phase**: *H* hidden layer output is calculated.
3. **Phase**: $\hat{\beta}$ output weights are calculated in accordance with. *Y* is the target feature.

2.2.2. Thyroid Diseases Diagnosis Method

Classification and recognition is an important issue in machine learning. Classification is the process of predicting of which group the objects are belongs to. In the study, Extreme Learning Machine (ELM) method was used in classification of thyroid disorders. Diagram belong to the study has been given in Figure 2; and ELM classification network model in Figure 3. The study composes of 4 blocks. Data set was collected in block 1. Data set were obtained from UCI machine learning repository. The data set compose of 215 examples and 5 features. The features acquired from block 2 have been normalized with min-max transformation. The features obtained from block 3 were classified with ELM. The last block includes the results of the classification.

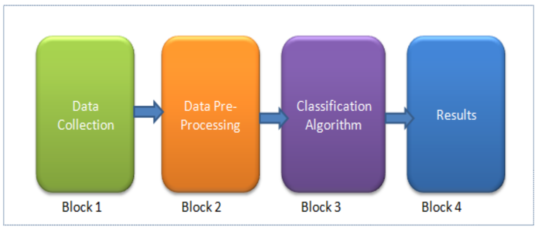

Figure 2. Diagram of classification duration for diagnosis of thyroid disorders.

3. RESULTS

In this study, the ELM was used in the diagnosis of thyroid disorders. Other studies in the literature carried on the same data set, and ELM classification accuracy results are given in Table 1.

Table 1. The classification accuracies obtained with ELM and other accuracies from literature for thyroid dataset.

Author	Method	Accuracy %
Özyilmaz and Yildirim (2002)	MLP with bp (3xFC)	86.33
Özyilmaz and Yildirim (2002)	RBF (3xFC)	79.08
Özyilmaz and Yildirim (2002)	CSFNN (3 x FC)	91.14
K.Polat et al., (2007)	AIRS (10xFC)	81.00
K.Polat et al., (2007)	AIRS with Fuzzy weighted pre-processing (10 x FC)	85.00
Paavo and Pasi (2008)	Level set classifier	96.44
Pasi (2004)	Linear Discriminant Analysis	81.34
Esin et al., (2011)	A hybrid model based PCA-SVM (10xFC)	93.77
Esin et al. (2011)	Model based GDA and WSVM	91.86
This study (2014)	**ELM (70-30% training-test)**	**96.79**

It is seen in Table 1 that the ELM provides an advantage in the diagnosis of thyroid gland disorders. Classification results and process durations obtained through ELM for data sets of various trainings and test rates are shown in Table 2. Results in Table 2 were come by using 30 neurons with sigmoid activation functions in hidden layer of network.

Table 2. Classification accuracies and durations for six train-test data partitions.

Model (training-test)	Classification accuracy of Training set (%)	Classification accuracy of Test Set (%)	Average of Classification Accuracy (%)	Training Time (sec)	Test Time (sec)
30-70%	98.46	90.69	94.575	<0.0001	<0.0001
40-60%	98.84	93.80	96.320	<0.0001	<0.0001
50-50%	97.22	94.39	95.805	<0.0001	<0.0001
60-40%	97.67	95.35	96.510	<0.0001	<0.0001
70-30%	96.69	96.88	96.785	<0.0156	<0.0001
80-20%	95.35	97.67	96.510	<0.0312	<0.0001

When looked at Table 2, high classification rates obtained through ELM for all training sets can be seen. Of six different sets, the highest classification accuracy for 70-30% part belongs to training test with 96.785%. The lowest classification accuracy for 30-70% data set is 94.575%. More, it can be seen in Table 2 that classification process was completed in a very short time. Training and test accuracy rates in terms of neuron number in hidden layer for 50-50% training set are given in Figure 3 and 4 while durations of classification are given in Figure 5 and 6.

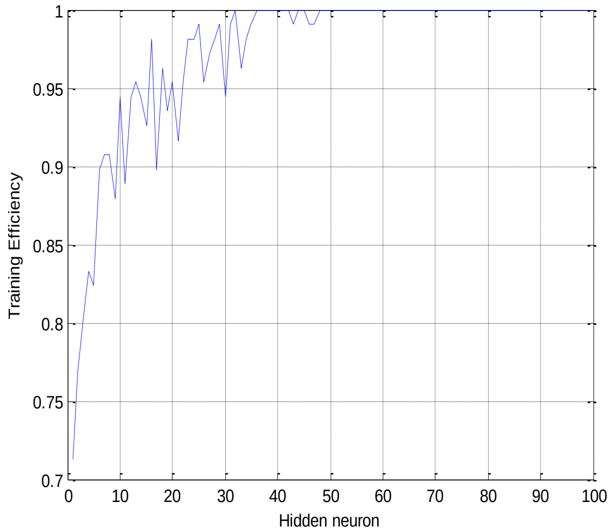

Figure 3. Training accuracies according to hidden neurons in hidden layer.

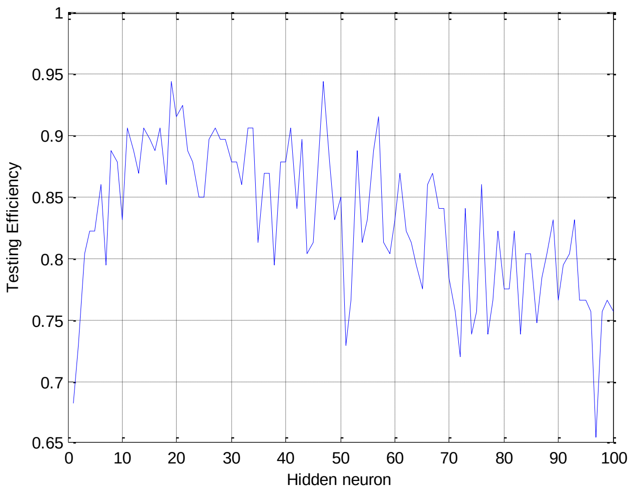

Figure 4. Testing accuracies according to hidden neurons in hidden layer.

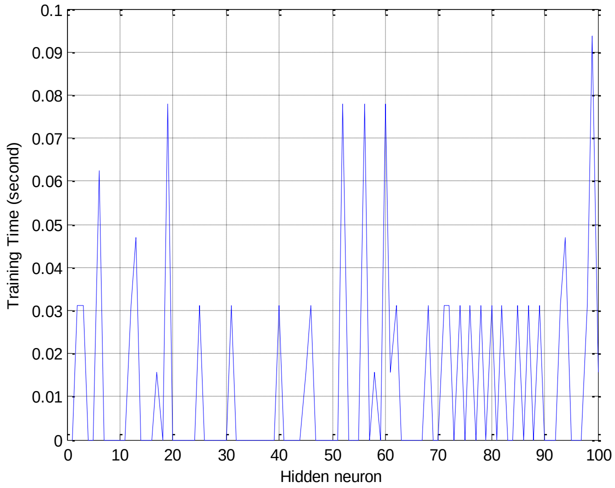

Figure 5. Training duration of training test according to hidden neurons in hidden layer.

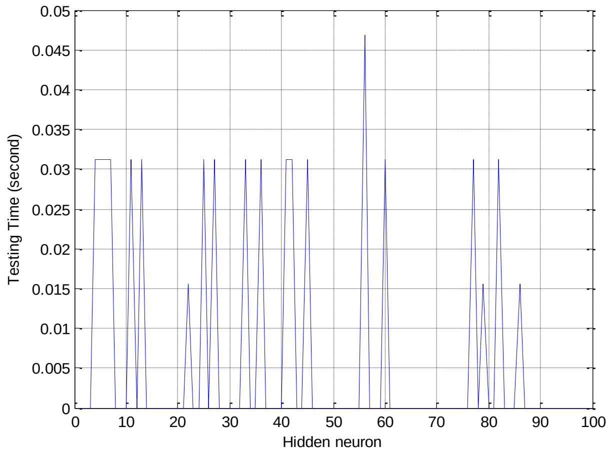

Figure 6. Classification durations of test data set according to hidden neurons in hidden layer.

Classification accuracy results obtained with respect to neuron number in hidden layer of 50-50% training test rates are given in Figure 7 and 8.

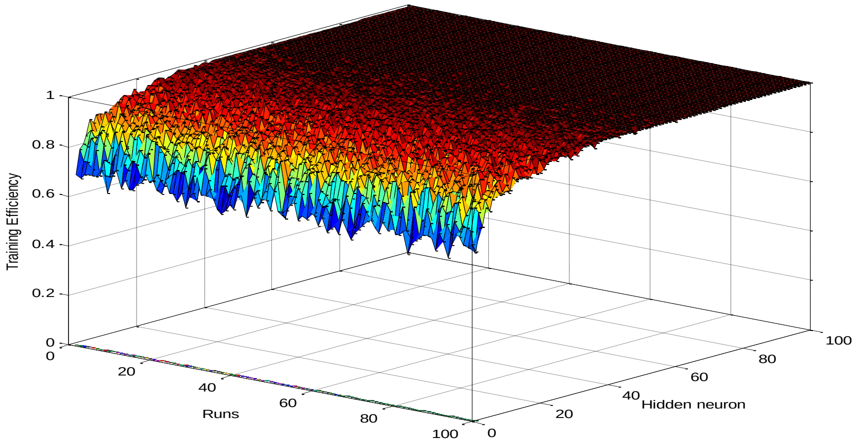

Figure 7. Training accuracy variation with respect to hidden neurons.

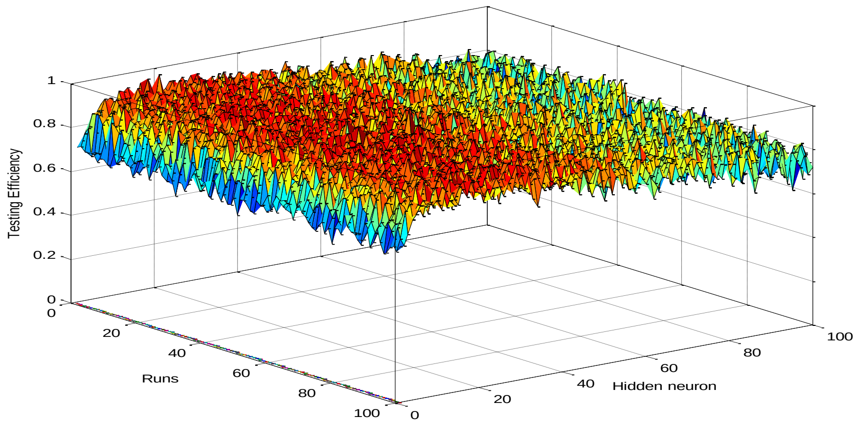

Figure 8. Testing accuracy variation with respect to hidden neurons.

As can be seen in Figure 7 and 8, accuracy rate increases with respect to neuron number in the hidden layer. It has been seen that accuracy rate for training data set remained stable after 30 neurons while accuracy rate for test data decreased. The highest accuracy rates were obtained when neuron number was in the range of 30-40. Accuracy results and durations of process obtained through various activation functions in the hidden layer for 50-50% training-test data set are stated in Table 3.

Table 3. Classification accuracy rates in accordance with various activation functions.

Activation Function	Training Performance (%)	Test Performance (%)	Training process Duration (sec)	Test Process Duration (sec)
Sigmoid	97.22	94.39	<0.0001	<0.0001
Sine	97.22	93.46	<0.0001	<0.0001
Hardlim	91.59	91.59	<0.0001	<0.0001
Triangular Basis	98.15	93.46	<0.0001	<0.0001
Radial basis	97.22	94.39	<0.0001	<0.0001
Probit	98.15	93.46	<0.0001	<0.0001

The best classification results were obtained by using sigmoid activation functions in the hidden layer.

Both classification accuracy and classification speed of ELM were compared to other methods. For 50-50% training-test data set, ELM was compared to Multilayer Perceptron (MLP), Support Vector Machine (SVM), Naive Bayes (NB), Decision Rules (J48) and Decision Trees (PART). Classification accuracy results acquired by various machine learning methods are shown in Table 4

Table 4. Classification accuracy rates acquired by various machine learning methods

Model	**Accuracy (%)**	**Classification Duration (sec)**
MLP	94.40	0.83
SVM	84.17	0.10
J48	91.58	0.10
PART	92.52	0.10
ELM (with sig Transfer function)	95.80	<0.0001

As shown in Table 4, the highest accuracy in terms of both classification accuracy and classification duration was obtained through ELM.

4. CONCLUSION

All metabolic functions of human beings are controlled by thyroid hormones. Thyroid is needed throughout the whole life, and any disorder with it must not be underestimated. Today in the diagnosis of thyroid disorders, machine learning methods as well as clinical studies are used widespread. In this study Extreme Learning Machine (ELM) which is new and fast in classification of thyroid gland disorders was used. According to the occurred findings, the ELM, for the data set, has the upper hand in terms of both classification accuracy and speed when compared to other machine learning methods. By using thyroid data set taken from UCI machine learning repository, 96.785% classification accuracy for 70-30% training-test partitions was obtained. ELM classification process was completed under 0.0001 sec**.** Our study indicates that the ELM could be an important model especially in medical applications where the velocity is important.

REFERENCES

Chih-Lin C. , W. Nick Street, David A. Katz, "A Decision Support System for Cost-Effective Diagnosis". *Artificial Intelligence in Medicine*, 50, 149-161, 2010.

Esin D., Akif D., Derya A., "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, 2011.

Handoko, S. D., Keong, K. C., Ong, Y. S., Zhang G.L., Brusic V, "Extreme Learning Machine for Predicting HLA-Peptide Binding", *Lecture Notes in Computer*, 3973, 716–721, 2006.

Hai-Jun R.,Yew-Soon O., Ah-Hwee T., Zexuan Zhu, "A Fast Pruned-Extreme Learning Machine for Classification Problem", *Neurocomputing*, 72, 359-366, 2008.

Halife K., Seral O., Arslan ,A., Gunes, S., "Medical Application of Information Gain Based Artificial Immune Recognition System (AIRS): Diagnosis of Thyroid Disease", *Expert Systems with Applications*, 36, 3086-3092, 2009.

Hoshi, K., Kawakami, J., Kumagai, M., Kasahara, S., Nisimura, N., Nakamura, H., "An Analysis of Thyroid Function Diagnosis using Bayesian-type and SOM-Type Neural Networks", *Chemical and Pharmaceutical Bulletin*, 53, 1570-1574, 2005.

Huang, G.B., Zhu, Q.Y., Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 70 (1-3), 489-501, 2006.

Kaya, Y., "A New Intelligent Classifier for Breast Cancer Diagnosis Based on Rough Set and Extreme Learning Machine: RS+ELM", *Turkish Journal of Electrical Engineering and Computer Sciences,* 21:2079-2091, 2013.

Kaya, Y., Uyar, M., "A Hybrid Decision Support System Based on Rough Set and Extreme Learning Machine for Diagnosis of Hepatitis Disease", *Applied Soft Computing Journal*, 13,3429- 3438, 2013.

Kaya, Y., Kayci, L., Tekin, R., Ertugrul, Ö.F., "Evaluation of Texture Features for Automatic Detecting Butterfly Species using Extreme Learning Machine", *Journal of Experimental & Theoretical Artificial Intelligence*, DOI:10.1080/0952813X.2013.861875, 2014.

Keles, A., Keles, A. "ESTDD: Expert System for Thyroid Diseases Diagnosis", *Expert Systems with Applications*, 34:1, 242-246, 2008.

Liang, N., Y., Saratchandran,P., Huang G., P., Sundararajan, N., "Classification of Mental Tasks From Eeg Signals using Extreme Learning Machine", *International. Journal of Neural Systems*, 16 (1), 29-38, 2006.

Ozyılmaz, L. Yıldırım, T., "Diagnosis of Thyroid Disease using Artificial Neural Network Methods", *In Proceedings of The 9th International Conference on Neural Information Processing* (ICONIP'02), 4, 2033-2036, 2002.

Paavo K., Pasi L., "Classification Method using Fuzzy Level Set Subgrouping", *Expert Systems with Applications*, 34, 859-865, 2008.

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.

Polat, K., Sahan, S., Gunes, S., "A Novel Hybrid Method Based on Artificial Immune Recognition System (AIRS) with Fuzzy Weighted Pre-processing for Thyroid Disease Diagnosis", *Expert Systems with Applications*, 32, 1141-1147, 2007.

Shu-Kay N., Geoffrey J. McLachlan. "Extension of Mixture-of-experts Networks for Binary Classification of Hierarchical Data", *Artificial Intelligence in Medicine*. 41, 57-67, 2007.

Suresh S., Saraswathi S., Sundararajan N., "Performance Enhancement of Extreme Learning Machine for Multi-category sparse Data Classification Problems", *Engineering Applications of Artificial Intelligence*, 23, 1149-1157, 2010.

Temurtas , T., "A Comparative Study on Thyroid Disease Diagnosis using Neural Networks", *Expert Systems with Applications*, 36, 944-949, 2009.

UCI Repository of Machine Learning Databases, University of California at Irvine, Department of Computer Science. <http://www.ics.uci. edu/_mlearn /databases/thyroid-disease/new-thyroid.data> Last Accessed 31.12.2011.

Yuan Q., Weidong Z., Shufang L., Dongmei C. "Epileptic EEG Classification Based on Extreme Learning Machine and Nonlinear Features", *Epilepsy Research*. 96, 29-36, 2011.

Zhang, G., Berardi, L. V., "An Investigation of Neural Networks in Thyroid Function Diagnosis", *Health Care Management Science*, 29–37, 1998.

Zong, W. W.,Huang G. B., "Face Recognition Based on Extreme Learning Machine", *Neurocomputing*, 74, 2541–2551, 2006.