

Received: 2018.04.11
Accepted: 2018.05.02
Published: 2018.09.17

Extreme Learning Machine (ELM)-Based Classification of Benign and Malignant Cells in Breast Cancer

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEFG **Abdullah Toprak**

Department of Biomedical Engineering, Engineering Faculty, Dicle University, Diyarbakir, Turkey

Corresponding Author: Abdullah Toprak, e-mail: atoprak@dicle.edu.tr

Source of support: Departmental sources

Background: Breast cancer is one of the most common cancer types in the world and is a serious threat to health. This type of cancer is complex; it is a hereditary disease and does not result from a single cause. The diagnosis of cancer starts with a biopsy. Various methods are used to detect and recognize cancer cells, from microscopic images and mammography to ultrasonography and magnetic resonance images (MRI).


Material/Methods: Detection and characterization of benign and malignant cells by image-processing-based segmentation for breast cancer diagnosis is important for early diagnosis. In the present study, Extreme Learning Machine (ELM) classification was performed for 9 features based on image segmentation in the Breast Cancer Wisconsin (Diagnostic) data set in the UC Irvine Machine Learning Repository database.

Results: The results obtained with the developed method were compared with the results of other machine learning methods (Naive Bayes, Support Vector Machine, and Artificial Neural Network) and it showed the highest performance, with a result of 98.99%.

Conclusions: It was found that both accuracy and speed were good. We present a method that can be applied in cell morphology detection and classification in automated systems that classify by computer-aided mammogram image features.

MeSH Keywords: **Breast Neoplasms • Classification • Diagnosis**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/910520>

 2420

 6

 5

 25



Background

Breast cancer is the second most common cancer in the world and is a serious threat to health. One out of 8 females have a high rate of having breast cancer at any age, especially after 40 years old. The best way to protect against breast cancer risk is an early diagnosis. New computer-assisted methods for breast cancer diagnosis make it possible to diagnose it faster and in a different way. Early detection of breast cancer can increase the patient's chances of survival. Breast cancer is a heterogeneous disease with different clinical outcomes. Moreover, tumor responses with different biological traits are followed by a long and challenging process [1–6]. Breast cancer is diagnosed by biopsy. A biopsy is a laboratory procedure to detect cancer and is performed by a pathologist. The pathologist collects tissue samples from breast regions. There are various techniques for collecting samples of breast tissue. These techniques include fine-needle aspiration, core-needle aspiration, core-needle biopsy, vacuum-assisted biopsy, and surgical biopsy. These cancer tissues are then analyzed by using a microscope. Images obtained from the microscope are also subjected to histopathological imaging. The pathologist analyzes the histopathological images and classifies them as cancerous or cancer-free images [7].

Previous studies on this topic have been performed. Zehra et al. [5] identified masses in mammograms and classified them as benign or malignant, showing that the surroundings of malignant tumors are more irregular and the size value is within certain limits. William et al. [6] estimated the classification of breast mass diagnosis based on fine-needle aspirates (FNA) by using digital image analysis and machine learning method; the predicted diagnostic accuracy was 97% and the actual diagnostic accuracy was 100% in 118 new samples. Mu and Nandi [4] used an automated classification methodology to analyze breast masses with a fine-needle aspirates (FNA) for diagnosis of malignancy and to characterize the features. Kowal et al. [11] proposed an approach for an automatic classification of images based on determining the kernel regions in the images and then using these regions as classifiers. Two-stage segmentation was applied to the images. In the first step, foreground and background segmentation was performed on images using the adaptive threshold. These images included the nuclei, red blood cells, and other features. In the second stage, the nucleus regions were separated from the blood cells and other features. Finally, the kernel regions were represented by different properties, and these properties are used as inputs for the classifiers. Using 3 classifiers (K-nearest neighbors, Naive Bayes, and decision trees), they obtained classification accuracy of 96–100% in 500 sample images. Filipczuk et al. [12] used computer-assisted breast cancer detection and proposed an approach to identify kernel areas and separate them into regions. The kernel areas were

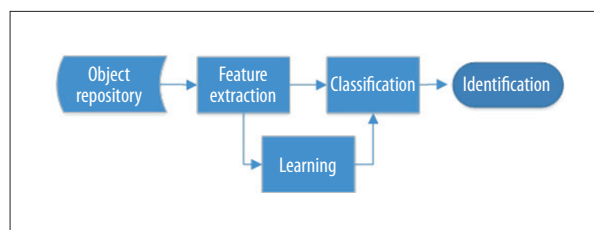


Figure 1. Tumor detection and classification system in breast cancer.

determined using the circular Hough transformation, as well as Kerbyson and Atherton methods. Then, the regions separated by the divisions were used as inputs to the classifiers. These classifiers (K-nearest neighbors, Naive Bayes, and Support Vector Machines) were found to have 98.51% classification accuracy in 737 microscopic images of fine-needle biopsies using these classifiers.

Since digital imaging has become an integral part of research, computer-aided assessment using advanced image analysis has become an important part of many research projects. Model recognition is a scientific research process used to determine the system design models by analyzing the data. The recognition system has emerged with a large accumulation of computer data [8].

The purpose of image processing is to increase the accuracy of computer control and verification of the data by activating many perceptions with large categories under different conditions. In the detection and diagnosis of breast cancer, mammography, ultrasonography, and magnetic resonance (MR) images are image-processing-based methods used to provide information [9]. Tumor recognition in the area examined with MR images is usually composed of 2 parts. The first part is the process of selecting the features required for recognition of the tumor or the sizes to be measured. This process is called feature extraction and it is the system feature extractor that performs this operation. Each feature used in tumor detection or the size to be measured is a real number that gives the measurement result. At the beginning of the determining the factors affecting the success of the classification, the selected features should represent all the tumor being sought. The second part of the tumor recognition is classified as benign and malignant, using the obtained feature vectors. In Figure 1, a diagram of tumor recognition system in breast cancer is presented.

In the present study, Extreme Learning Machine (ELM)-based on classification was performed on 9 features based on image segmentation in the Breast Cancer Wisconsin (Diagnostic) data set in the UC Irvine Machine Learning Repository database. The features and innovations of this study are the classification of tumors based on image segmentation, Extreme

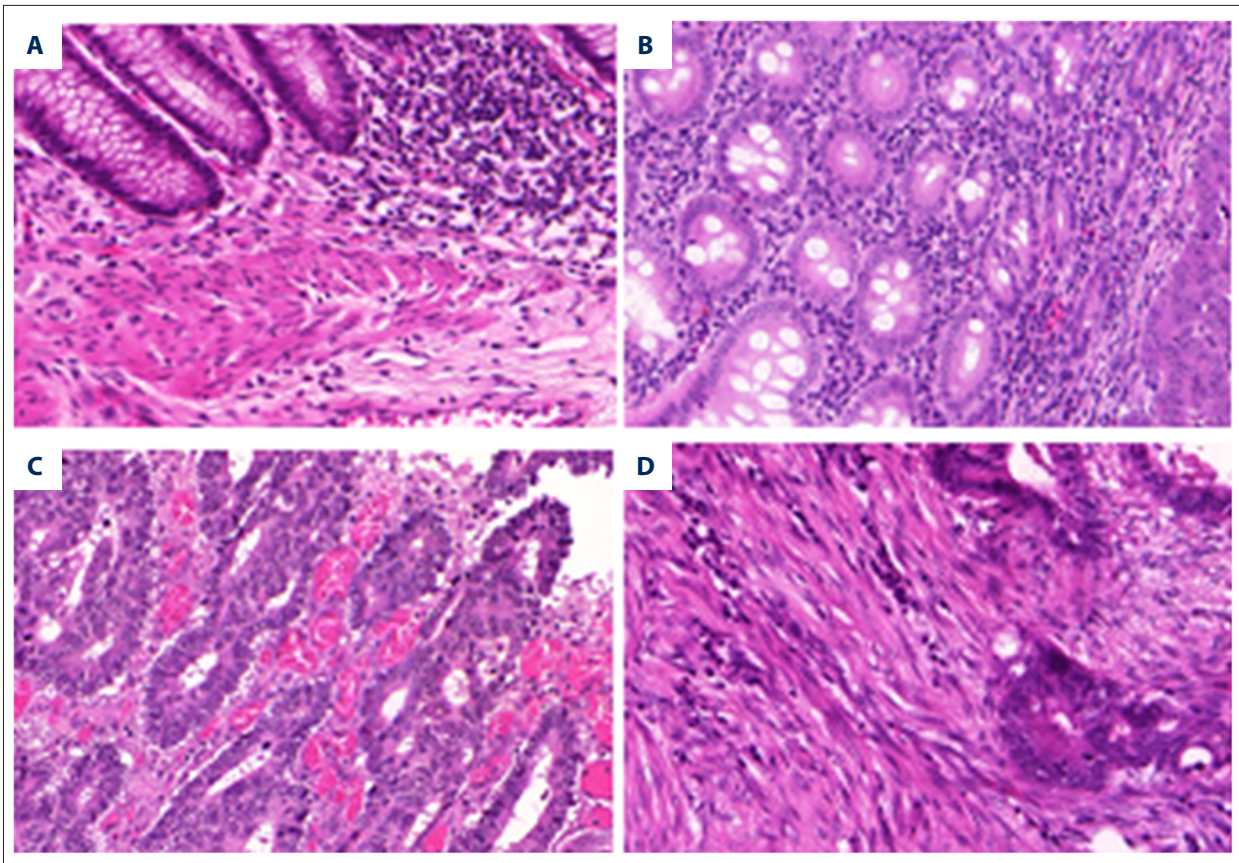


Figure 2. Benign and malignant cell form. (A) Benign, adenomatous. (B) Benign, healthy. (C) Malignant, poorly differentiated. (D) Malignant, moderately differentiated.

Learning Machine (ELM) classification, and regression methods. Figure 2 shows images of benign and malignant cell forms [10].

Material and Methods

The database used in this research is the Breast Cancer database from the UCI Machine Learning Repository, created by Dr William H Wolberg at the University Hospital of Wisconsin, Madison, containing 699 data points belonging to 2 categories – benign and malign (malignant) – in which 458 of the samples belong to the malignant class of the benign class 241 and 9 features based on image segmentation were removed from the data set. Input and output for classification with properties and value ranges are shown in Table 1 [6].

The approach for solving a problem using machine learning techniques are [13]:

- Definition of the problem
- Data set
- Modeling
- Model performance evaluation

Table 1. Table of properties.

Properties	Value range	Category
Clump thickness	1–10	2 – Benign 4 – Malign
Uniformity of cell size	1–10	
Uniformity of cell shape	1–10	
Marginal adhesion	1–10	
Single epithelial cell size	1–10	
Bare nuclei	1–10	
Bland chromatin	1–10	
Normal nucleoli	1–10	
Mitoses	1–10	

Definition of the problem

We3 tried to solve the problem of classification of benign and malignant cells based on image segmentation for breast cancer detection and classification using the machine learning method.

Data set

The database used in this research is the Breast Cancer database from UCI Machine Learning Repository, created by Dr William H Wolberg at the University Hospital of Wisconsin, Madison.

Modeling

Once the data are ready to be processed, the modeling phase starts for the learning algorithm. The model is basically the architecturalization of the need for output defined in accordance with the attributes of the task.

In this study, the ELM model was used as a new approach to evaluate data for the classification of benign and malignant cells based on image segmentation for the detection of breast cancer. Information on working algorithms of the models is presented in the subsections.

Extreme learning machine

Extreme Learning Machine (ELM) is a single hidden-layer feed-forward neural network (SLFN). The performance of the SLFN should be appropriate for the system to be modeled for data such as threshold value, weight, and activation function so that higher learning can be performed. In gradient-based learning approaches, all of these parameters are iteratively changed for the appropriate value. Therefore, due to the possibility of being attached to the slow and local minimum, the performance can produce low results. Unlike FNN, which is renewed on the basis of the gradient in the ELM Learning process, the output weights are analytically calculated while the input weights are randomly selected. In an analytic learning process, the success rate increases because the resolution time and the error value can seriously reduce the possibility of being fitted to a local minimum. ELM can be used to select a linear function to activate cells in the hidden layer, as well as to use non-linear (sigmoid and sinusoidal), non-derivatized, or intermittent activation functions [14–18]. Figure 3 shows the ELM architecture.

$$y(p) = \sum_{j=1}^m \beta_j g \left(\sum_{i=1}^n w_{ij} x_i + b_j \right) \tag{1}$$

β_j represents the weights between the input layer and the hidden layer and β_1 represents the weights between the output layer and the hidden layer. b_j is the threshold value of the neurons in the hidden layer, $g(\cdot)$ activation function. Equal input layer weights (w_{ij}) and bias (b_j) are randomly assigned. The activation function ($g(\cdot)$) is assigned at the beginning of the input layer neuron number (n) and hidden-layer neuron number (m). Now, based on this information, if the parameters known in equilibrium are combined and rearranged, the output layer becomes as in Equation 3.

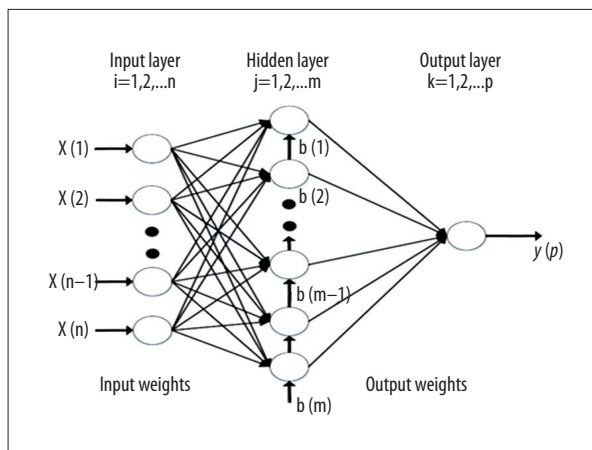


Figure 3. A single hidden-layer feed-forward neural network model.

$$H(w_{ij}, b_j, x_i) = \begin{bmatrix} g(w_{1,1}x_1 + b_1) & \dots & g(w_{1,m}x_m + b_m) \\ \vdots & \ddots & \vdots \\ g(w_{n,1}x_1 + b_1) & \dots & g(w_{n,m}x_m + b_m) \end{bmatrix} \tag{2}$$

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

In all training algorithm models, the goal is to minimize the error as much as possible. The error function of the output Y_p obtained by the actual output \hat{Y}_o value in ELM is $\sum_k^s (\hat{Y}_o - Y_p)$ (with “s”: number of training data) $\| \sum_k^s (\hat{Y}_o - Y_p)^2 \|$ can be minimized. For both of these functions, the output Y_p obtained by the actual output value Y_o must be equal to Y_p . When this equation is satisfied, the unknown parameter in Eq. The H matrix can be a matrix with a very low probability, meaning that the number of data in the training set is unlikely to be equal to the number of features that each data contains. Therefore, taking the inverse of $[H^-]$ and finding weights (β) will be a problem. To overcome this situation, Huang et al. [17] proposed using the generalized inverse Moore-Penrose matrix, which was developed to calculate approximate inverses of matrices that cannot be reversed, as in this problem. $\hat{\beta}$ is the output weight and H^+ is the generalized inverse Moore-Penrose matrix of H matrix. Accordingly, the output weights can be found by $\hat{\beta} = H^+$ [14,15].

Model performance evaluation methods and criteria

It is important to compare the performance values of machine learning algorithms with a measurable expression and to compare its performances. In this section, we have divided the data set, which is the first criterion affecting the performance of the model and the algorithm used, divided into training and test data, and the second is the identification of performance evaluating expressions.

➤ *First criterion:* There are various data partition performance evaluation methods such as hold-out and K-fold cross-validation in the literature [13, 19]. The following items should

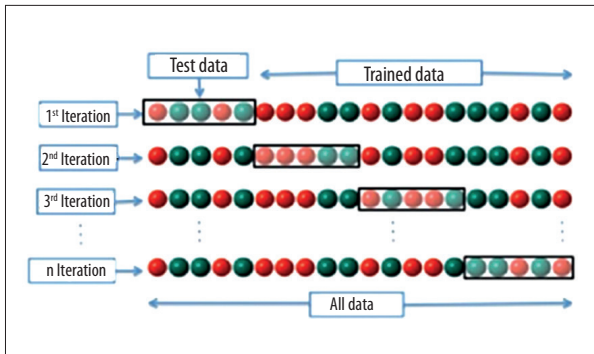


Figure 4. K-fold cross-validation performance evaluation method.

be taken into consideration while distinguishing the data set as training and testing:

- The number of samples in the training dataset should be more than the samples in the tested dataset.
- It is necessary to randomly distribute samples at the distinction of training and test data set.
- During the division of the data set into training and test data sets, the target class must include the target data in the distribution of the training and test data sets.

In the K-fold method, the data set is divided into 3 parts – training, verification, and test data – by 3 steps – separation, model selection, and performance status – which are made at the same time. For the ELM used in this study, the K-fold cross-validation method was chosen as the data set divided into training and test data, as shown in Figure 4.

➤ *Second criterion:* It is necessary to express the performance of the proposed solution for a probing given in machine learning algorithms, or to express how well the algorithm learns. Different evaluation criteria have been developed for this.

In order to perform the performance evaluation of the classifier models to be applied to the dataset, we used criteria of accuracy, sensitivity, determinism, precision, and f-measure, which are explained below by creating an error matrix [20–25] (Table 2).

Table 2. Error matrix.

		Real value (detection)		
		Real positive (yes)	Real negative (no)	Total (real)
Predicted value	Predicted positive (yes)	True positive (T_{pos})	False positive (F_{pos})	Tot_{pos}
	Predicted negative (no)	False negative (T_{neg})	True negative (F_{neg})	Tot_{neg}
	Predicted total	pos	neg	Tot

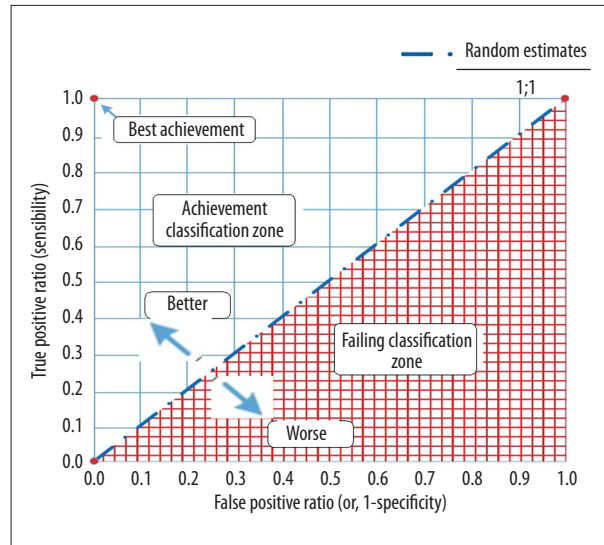


Figure 5. Representation of ROC in space.

Accuracy value is measured as the ratio of all data in the data set of the data correctly guessed by the algorithm with correct detection.

$$Accuracy = \frac{T_{pos} + T_{neg}}{Tot} \quad (4)$$

Sensitivity value can be measured as the ratio of true real positive to all true and false real positive.

$$Sensitivity = \frac{T_{pos}}{T_{pos} + F_{pos}} \quad (5)$$

Specificity value can be measured as the ratio of true real negative to all real negative.

$$Specificity = \frac{T_{neg}}{T_{neg} + F_{neg}} \quad (6)$$

Precision is the rate of correct estimation and can be measured as the ratio to all real and predicted positives.

$$Precision = \frac{T_{pos}}{T_{pos} + F_{pos}} \quad (7)$$

The F-measure is a harmonized mean of the sensitivity and precision measures.

$$F - measure = \frac{2}{\frac{1}{Sensitivity} + \frac{1}{Precision}} \quad (8)$$

Table 3. ELM classification results.

Test performance/ accuracy performance	Number of cells in hidden layer	Activation function
98.99%	1000	Linear (lin)

Table 4. ELM parameters.

Number of cells in hidden layer 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
Activation function: sigmoid (sig), sinus (sin), hard limit (hardlim), triangular basis (tribas), radial basis (radbas) ve linear (lin)

The receiver operating characteristics (ROC) graph is a frequently used graph that summarizes the performance of the curve classifier over all possible thresholds. When you change the threshold of a particular class observation assignment, you plot the true positive rate or sensitivity value against the false positive rate or 1-value of determinism (Figure 5).

Results

The results of benign and malign classification for the data set by ELM method are shown in Table 3. The K-fold diagonal data segmentation method is used for data set partition. During the process, the ELM cell count and activation function in the hidden layer (Table 4) were evaluated to be maximum (close to 1) from the results obtained using the Dene-Yanil method. When the value is determined by the object property found in the data set, the ELM can be said to be 98.99% or 98.99 in every 100 datasets based on these values. ELM showed 100% performance in network training and 98.99% performance in training. These results were obtained with MATLAB 2106a software and a laptop computer with an Intel i7-6500 CPU and 16 GB of RAM. This performance was recorded as 0.0078 s at the training time of the network and 0.0052 s at the testing time. When these results are obtained, the number of cells in the hidden layer for ELM is 1000 and the activation function is linear.

The performance of the ELM method and other machine learning methods – Support Vector Machine (SVM) and Naive Bayes (NB) – are given in Table 5, showing that ELM is superior to other methods in performance and speed. The performance metric values obtained by applying other machine learning classifiers to the test data are shown in Table 6.

Table 5. Performance of different classification methods.

Method	Training result	Test/ accuracy result	Test time
ELM	100%	100.00%	0.0052 s
SVM	100%	96.85%	0.06 s
NB	100%	95.99%	0.04 s

Table 6. Performance metric values of alternative method classifiers.

	Performance metrics			
	Sensitivity	Specificity	Precision	F-measure
Support vector	0.98	0.94	0.96	0.97
Naive Bayes	0.98	0.91	0.95	0.96

Discussions

In this study, input and output parameters for feature ELM classifier based on image segmentation of mammogram were determined. Benign and malignant classification was performed with ELM. Performance and speed were measured and a comparison with other methods of classification was performed. The performances of the ELM method and other machine learning methods – Support Vector Machine (SVM) and Naive Bayes (NB) – allowed showing not only the results of ELM method, but also comparing it with other methods, and this helps pathologist to use and choose the fastest method with the best performance. This study can be continued in the future by comparing this classification method with others but taking into account other parameters of comparison or other methods that were not explored here.

Conclusions

We compared the results obtained with ELM with results obtained with other classifier (SVM and NB) methods, showing that accuracy and speed were good. We present a method that can be applied in cell morphology detection and classification in automated systems that classify by use of computer-aided mammogram image features. This gain in accuracy and speed for the classification of benign and malignant cells lead to more efficient breast cancer diagnosis.

References:

1. Seng WC, Mirisae SH: A new method for fruits recognition system. In *Electrical Engineering and Informatics 2009. ICEEI'09. International Conference on 2009 Aug 5; Vol. 1, 130–34*
2. Bassett LW, Hoyt AC, Oshiro T: Digital mammography: Clinical image evaluation. *Radiol Clin North Am*, 2010; 48(5): 903–15
3. Ku JL, Park SC, Kim KH et al: Establishment and characterization of seven human breast cancer cell lines including two triple-negative cell lines. *Int J Oncol*, 2013; 43(6): 2073–81
4. Mu T, Nandi AK: Breast cancer diagnosis from fine-needle aspiration using supervised compact hyperspheres and establishment of confidence of malignancy. In *Signal Processing Conference 2008 16th European 2008 Aug 25; 1–5*
5. Karhan Z, Ergen B, Cetintas D: Mass detection on mammograms with image processing techniques and benign-malignant distinction. In *Signal Processing and Communications Applications Conference (SIU) 2014 22nd 2014 Apr 23; 846–49*
6. Wolberg WH, Street WN, Heisey DM, Mangasarian OL: Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. *Arch Surg*, 1995; 130(5): 511–16
7. Kaymak S, Helwan A, Uzun D: Breast cancer image classification using artificial neural networks. *Procedia Computer Science*, 2017; 120: 126–31
8. Dedeoğlu Y, Töreyn BU, Gündükbay U, Çetin AE: Videoda Nesne Sınıflandırması için Siluet Tabanlı Yöntem. *Signal Processing and Communications Applications, 2006; 1–4* [in Turkish]
9. Domeniconi C, Peng J, Gunopulos D: An adaptive metric machine for pattern classification. In: *Advances in Neural Information Processing Systems*, 2001; 458–64
10. Sirinukunwattana K, Snead DR, Rajpoot NM: A stochastic polygons model for glandular structures in colon histology images. *IEEE Trans Med Imaging*, 2015; 34(11): 2366–78
11. Kowal M, Filipczuk P, Obuchowicz A et al: Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Comput Biol Med*, 2013; 43(10): 1563–72
12. Filipczuk P, Fevens T, Krzyzak A, Monczak R: Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE Trans Med Imaging*, 2013; 32(12): 2169–78
13. Ertam F: Kurumsal bilgisayar ağlarındaki trafik bilgisinin akıllı sistemler ile sınıflandırılması, Doktora Tezi, F.Ü. Fen Bilimleri Enstitüsü, Elazığ, 2016 [in Turkish]
14. Ertuğrul ÖF, Tağluk ME, Kaya Y, Tekin R: EMG signal classification by extreme learning machine. In: *Signal Processing and Communications Applications Conference (SIU)*, 2013 21st 2013 Apr 24; 1–4
15. Kaya, Y, Tekin R: Epileptik Nöbetlerin Tespiti için Aşırı Öğrenme Makinesi Tabanlı Uzman Bir Sistem. *International Journal of Informatics Technologies* 2012; 5(2) [in Turkish]
16. Wang G, Li P: Dynamic Adaboost ensemble extreme learning machine. In: *Advanced Computer Theory and Engineering (ICACTE)*, 2010 3rd International Conference on 2010 Aug 20; Vol. 3: V3–54
17. Huang GB, Zhu QY, Siew CK: Extreme learning machine: Theory and applications. *Neurocomputing*, 2006; 70(1–3): 489–501
18. Huang GB, Zhu QY, Siew CK: Extreme learning machine: A new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on 2004 Jul 25; Vol. 2: 985–90*
19. Subudhi BN, Ghosh S, Shiu SC, Ghosh A: Statistical feature bag based background subtraction for local change detection. *Information Sciences*, 2016; 366: 31–47
20. Japkowicz N, Shah M: *Evaluating learning algorithms: A classification perspective*. Cambridge University Press, 2011
21. Hamel L: Model assessment with ROC curves. In: *Encyclopedia of Data Warehousing and Mining, Second Edition*, 2009; 1316–23
22. Fawcett T: An introduction to ROC analysis. *Pattern Recognition Letters*, 2006; 27(8): 861–74.
23. Forman G: An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 2003: 1289–305
24. MATLAB. [cited 2018 Jan 15] Available from: URL: <http://www.mathworks.com/products/matlab>
25. WEKA. [cited 2018 Jan 15] Available from: URL: <http://www.cs.waikato.ac.nz/ml/weka>