



Proposing an Integrated Method based on Fuzzy Tuning and ICA Techniques to Identify the Most Influencing Features in Breast Cancer

Irvan Masoudiasl¹, Shaghayeh Vahdat², Somayeh Hessam^{2,*}, Shahaboddin Shamshirband^{3,4,**} and Hamid Alinejad-Rokny^{5,6}

¹Department of Healthcare Services Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

²Department of Health Services Administration, South Tehran Branch, Islamic Azad University, Tehran, Iran

³Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁴Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁵The Graduate School of Biomedical Engineering, UNSW Australia, 2052, Sydney, Australia

⁶School of Computer Science and Engineering, UNSW Australia, 2052, Sydney, Australia

*Corresponding author: Department of Health Services Administration, South Tehran Branch, Islamic Azad University, Tehran, Iran. Email: somayehh59@yahoo.com

**Corresponding author: Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam. Email: shahaboddin.shamshirband@tdtu.edu.vn

Received 2019 April 08; Accepted 2019 August 17.

Abstract

Background: Breast cancer is the most common cancer in women, which has not been completely cured yet. The traditional approaches have low accuracy for breast cancer detection. However, intelligent techniques have been recently used in medical research to distinguish infected individuals from healthy ones, accurately.

Objectives: In this study, we aim to develop an ensemble machine learning (ML) method to distinguish tumor samples from healthy samples robustly.

Methods: We used an Imperial Competitive Algorithm coupled with a Fuzzy System (ICA-Fuzzy-SR) to identify the most influencing features to recognize tumor samples. To evaluate the proposed method, we used the publicly available Wisconsin Breast Cancer Dataset (WBCD).

Results: Benchmarking with the current existing leading methods indicates that our proposed method achieves 95.45% prediction accuracy, which is 3% better than those reported in previous studies.

Conclusions: Such results achieve while our model is significantly faster than previously proposed models to solve this problem.

Keywords: Algorithms, Benchmarking, Breast Neoplasms, Fuzzy Tuning, ICA Feature Selection, Machine Learning, Sparse Representation, Wisconsin

1. Background

When abnormal cells grow up in breast tissue, the result is to create breast cancer. This type of cancer is common in developing and advanced countries. The changed lifestyle and hormone therapy are among the main reasons for breast cancer. Diagnosing breast cancer at an early stage helps initiation of less extensive treatment (1). Most of the diagnostic methods are based on expensive and time-consuming methods such as, surgery and mammography screening. In developed countries, mammography screening is a conventional method (2). However, there is a need for new cost-effective approaches for early diagnosis. Unfortunately, the typical strategy for detecting breast cancer has failed to achieve high accuracy in diagnosis management and inaccurate alert rates. Therefore, artificial intelligence (AI) and specifically, machine learn-

ing (ML) approaches are adopted to make smart decisions (3).

During the past decade, machine learning (ML) techniques have been widely employed to diagnose breast cancer for tumor feature extraction in order to increase diagnostic accuracy (4). Fuzzy logic is considered as a primary method in ML to tackle this problem. Most applications of this model are for the classification and feature extraction of tumors in this field of the study. This method has been employed for the classification of various tumors. Gerald Schaefer et al. (5) employed a fuzzy logic method for the classification of breast cancer using bilateral differences among right and left breast areas. Also, Chen et al. (6) implemented a fuzzy c-mean (FCM) method for the segmentation of breast lesion in MR-images. In general, the fuzzy logic method has been employed to decide on uncertain

information. However, there is a need for dealing with the cognitive uncertainties of the learning mechanism.

Artificial Neural Network (ANN), an advanced technique in cognitive science and machine learning, has also been widely applied as a classification technique in cancer diagnosis. Khan et al. (7) developed an ANN to classify cancer for multiple diagnosis groups in the presence of their gene expression specifications.

Another important aspect of breast cancer research is to select the most informative and relevant feature for the identification process. Feature selection algorithms help the diagnose algorithm for having input features/attributes with more critical information. In this study, we aim at using Imperial Competitive Algorithm (ICA) to select more relevant features. ICA is an optimization method in AI that performs neighborhood movements and benefits from less dependency on the initial solutions. It has a better convergence rate compared to other optimization algorithms. Recently, ICA has been used in optimization algorithms, dimensionality reduction, and feature selection in many research topics such as medical research (8), (9), and (10).

Weighting the selected features or important attributes/features is a general method for highlighting the input features in the diagnostic/detection applications. The fuzzy system is one of the weighting methods that is employed for automatic weighing on uncertain features (11). The fuzzy system is proved to be practical in terms of automated decision making to assign a suitable weight to element. On the other hand, the manual weighting is proved to be unable to select the most informative features in comparison to fuzzy weighting (5).

The main objectives of this study are as follows:

- Investigating the state of art of the ML methods applied in breast cancer diagnosis.
- Performing the ICA (Imperialist competitive algorithm) feature selection method.
- Fuzzy tuning of features.
- Performing a sparse representation method for classifying the breast cancer data.
- Analyzing the results of the proposed method.

To achieve the objectives, we first select the best features through ICA and then use fuzzy rules to tune their weights. We then perform the sparse representation algorithm to classify the outcome of the dataset. The sparse representation for classification (SRC) is employed when specific data transformation techniques are applied. In other words, the main advantage of the sparse algorithm is ignoring the Euclidean distance between the samples when learning the wrong description of a test instance in the given dictionary (12).

The rest of this article is organized as follows. Section

2 presents a brief review of previous studies (literature review), and Section 3 presents the applied methodology. The proposed methods of applying ICA feature selection, Fuzzy tuning of feature and sparse representation are described in Section 4. The experimental results are presented in Section 5, and finally, the conclusion is provided in Section 6.

2. Literature Review

Breast cancer, as one of the most common cancers, has a high mortality rate in both developing and developed countries. From the estimated 1600 cured cases of breast cancer globally in 2012, 794 cases accrued in the more developed countries compared to 883 cases in the developing countries (13).

Image processing and mammograms screening are two most common breast cancer diagnosis approaches. Computer-Aided Design (CAD) Mammography is aimed to design a system with the ability to read mammograms independently in the presence of Convolutional Neural Network (CNN) (14).

With the advancement of computational methods, machine learning and data science are now considered as other significant contributors to medical science. For example, Rule-Based Fuzzy Cognitive Map (RBFCM) is considered as an important approach to classify the potential breast cancer threatening factors using human experiences and knowledge (15).

Williams et al. (16) employed the C4.5 (j48 is C4.5 version of weka) decision trees and the naïve Bayes' algorithms to predict breast cancer risks in Nigerian patients. In this study, C4.5 demonstrated better performance compared to naïve Bayes.

In another study (17), Random Forest was demonstrated to have the best classification performance for the majority of the tested group, while in the case of masses texture, Naive Bayes had the best performance.

In a different study conducted by Diz et al. (17), a data mining method has been presented for breast cancer diagnosis in the presence of two breast cancer datasets. The proposed algorithm was developed on WEKA to classify the texture features.

One of the primary time consuming and challenging tasks in this area is to filter all the information pertinent to support the clinical disease diagnosis. For this reason, a hybrid K-means and support vector machine (K-SVM) method was developed in (18) to extract useful information and diagnose the tumor. Using this model, they were able to significantly improve breast cancer detection (18).

Artificial immune recognition system (AIRS) has been successfully employed for diagnosing various diseases. In the study by Saybani et al. (19), AIRS was applied to classify

a tuberculosis disease. The proposed approach accurately classified the tuberculosis cells. Also, in another study by Saybani et al. (20), a hybrid AIRS-SVM, fuzzy logic, and real tournament selection mechanism were presented to detect the specified features of cancer cells in breast tissue. The results indicated a promising high-performance precision.

To the best of our knowledge, ICA technique provides on the highest accuracies among the other techniques for classification purposes in breast cancer research and feature extraction. Also, it can be comfortably merged with Computational Intelligence (CI) techniques to make hybrid methods.

Yaghoobi et al. (8) diagnosed breast cancer by Gray Level Co-occurrence Matrix (GLCM) and cumulative histogram features. They used the texture attributes from GLCM and presented the ICA feature selection method, decision tree, and neural network for feature selection. They showed that the ICA feature selection had better performance in terms of accuracy rate in diagnosing breast cancer. In (9), the authors proposed an algorithm that used a fuzzy expert system and the ICA as data mining methods to diagnose coronary artery disease (CAD). In this study, a fuzzy system and a decision tree were used to make correct rules. Also, the ICA was employed to modify fuzzy membership functions. Another application of ICA in medical research was for selecting the parameters of Gabor filters in retinal vessel segmentation (10). In this study, ICA reduced the parameters of the Gabor filter from 180 to 20 with the maximum accuracy for parameter selection. There are various dimensionality reducing and feature selection algorithms, like PCA and Fisher (21). However, the results have showed that the ICA feature selection algorithm after iterative epoch for selecting beneficial features has better performance in the low-dimensional dataset compared to PCA and LDA (22).

Similar to ICA and Fuzzy system, the sparse representation for classification has extensive applications in medical research. Al-Shaikhli et al. (23) used a sparse representation for local and global image information in 3D liver segmentation. They used the K-SVD method for the learning dataset. In (12), the authors applied the sparse representation based on the texture classification for CT lung images. They explored characterizing lung textures with sparse decomposition by different regularization techniques. Li et al. (24), proposed contour sparse representation methods and organ location determination for organ segmentation. They investigated the liver, kidney, and spleen.

Later on Jiang et al. (25), examined an images fusion strategy in medical images. They decomposed medical source images into low-frequency (LF) and high-frequency layers. In this study, the WBCD benchmark was extracted

from the UCI (University of California Irvine) machine learning repository and used for experimentations (Frank and Asuncion 2010) make consistent referencing. WBCD dataset included 699 samples which were collected from needle aspirates from human breast cancer tissue. This dataset consists of nine features from fine-needle aspirates in the form of an integer value between 1 and 10. The regular characteristics are Normal Nuclei, Uniformity of Cell Size, Shape Single Epithelial Cell Size, Marginal Adhesion, Uniformity of Cell, Mitoses, Bland Chromatin, Clump Thickness, and Bare Nuclei. In the employed dataset, 65.5% of the samples are benign, and 34.5% of samples corresponds to malignant class (26).

3. Research Methodology

Figure 1 shows the architecture of the proposed method for the detection of breast cancer based on the ICA feature selection, Fuzzy tuning, and Sparse Representation.

The steps of the proposed method in Figure 1 are described as follow:

1. Step 1: Splitting the dataset for train and test with leave one out cross-validation (LOOCV) or K-Fold cross-validation.
2. Step 2: Selecting the effective features on train-data via the ICA feature selection algorithm.
3. Step 3: Weighting the features by Fuzzy tune on features;
 - 3.1. Weighting features of selected indexes on test-data.
 - 3.2. Weighting features of selected indexes on train-data.

This weighting is performed via Fuzzy tuning of the feature.

4. Step 4: Performing sparse representation on dataset samples. This step represents the discriminative sparse coefficient for each sample.

5. Step 5: Classification step using Euclidean distance for discriminating the new sparse representation of dataset samples form a test sample of the dataset.

6. Step 6: Evaluation of the proposed method with several measurements criteria such as Accuracy, RMSE, R, R^2 , Sensitivity, and Specificity.

4. Proposed Method

4.1. ICA Feature Selection

There are various dimensionality reduction and feature selection algorithms such as Principal Component Analysis (PCA), Fisher, etc. (21). However, the results show that the ICA feature selection algorithm has been shown to

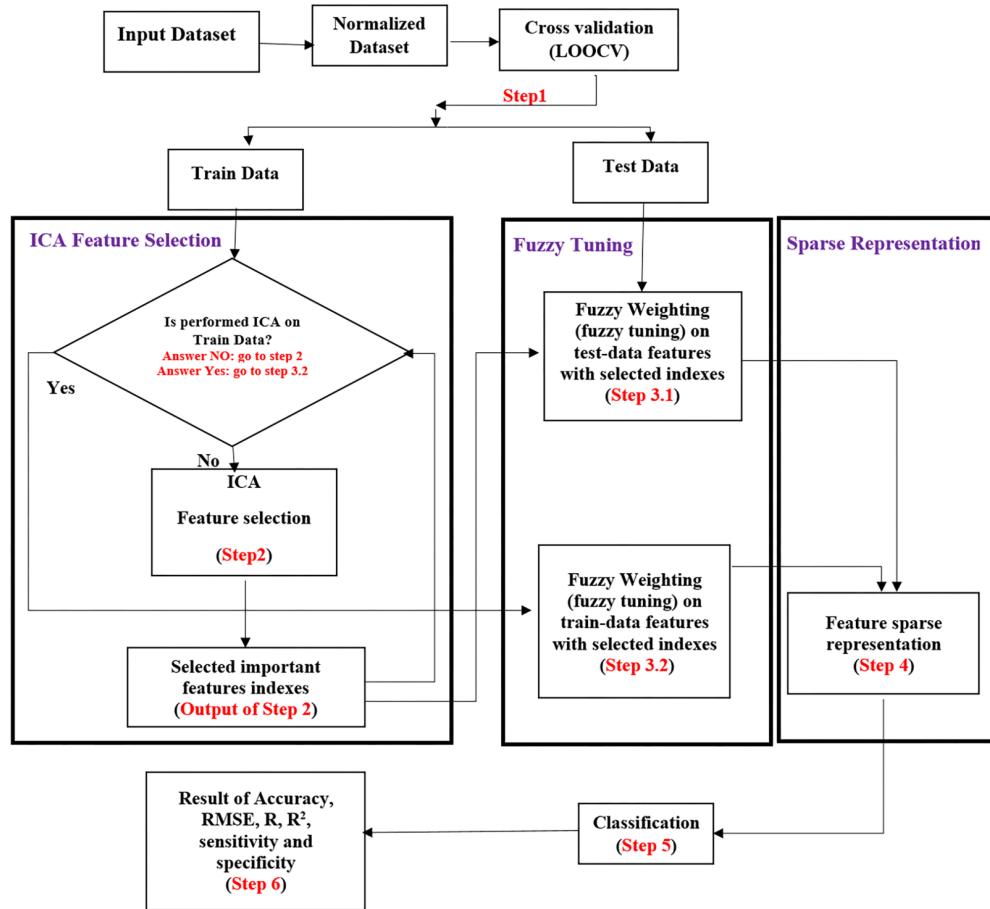


Figure 1. The general architecture of our proposed method for Detection of Breast Cancer

be better in a low-dimension dataset (22). ICA has been introduced, and its mechanism is explained in detail in (27). In this study, the ICA feature selection was adopted to a BC dataset to choose the most influencing features. The mechanism of the ICA is summarized in Algorithm 1 (Box 1).

According to the ICA procedure presented in Table 1, the best empire in ICA has the best cost, and its selected position is regarded as the best feature among all features. For example, according to the regularization parameters in Table 1, four elements among the nine features were selected in this study.

Table 1 shows the results of the ICA feature selection. The total number of populations was 200, and we selected 20 IMP randomly among the total population. One could find the weakest empire and give its colony to the winner empire. The total number of selected IMP was 17 for running the algorithm. The best cost was realized for IMP number 8. The best position of imperialist was [1,1,0,0,0,0,1,0,1]

where the value of 1 is the best feature selected by ICA, and the value of 0 is the unselected feature in the dataset.

4.2. Tuning of the Attributes Using the Fuzzy System

The ICA feature selection method selected four features, including: F1 (normal nuclei), F2 (uniformity of cell size), F7 (bland chromatin), and F9 (bare nuclei candidate as the best features). To make more differences between the selected features, we utilized the fuzzy system to tune the weight of each attribute. The fuzzy system produced a weight as an output of the fuzzy rule-based system (28). The mechanism of how fuzzy system perform is explained in detail in (29). This weight was calculated through the inference fuzzy rule-based system between F1, F2, F7, and F9 as the inputs of the fuzzy system.

Before applying the fuzzy system, we analyzed the data statistically to make sure about the range of each input to assign the proper membership function. In this case, the

Box 1. Imperialist Competitive Algorithm (ICA)

Procedure

1. Generate an initial population
2. Calculate the initial population's cost function
3. Sort initial population based on cost function values
4. Select imperialist state
5. Divide colonies among imperialists based on their cost function
6. Check the cost of all colonies in each empire
- 7. Assimilation:**
8. If there is a colony which has a lower cost than its imperialist, then
 9. Exchange the position of the colony and the imperialist
10. End if
- 11. Revolution:**
12. Update the position of the i^{th} empire
13. Calculate the empire's total cost
14. Find the weakest empire
15. Give one of its colony to the winner empire
16. Check the number of colonies in each empire
17. If there is an empire without a colony, then
 - 18. Remove the empire and give its imperialist to the best empire**
19. End if

statistical analysis of BC sample features in two classes (benign and malignant) showed that the value of features was different in each class (see Figure 2). For example, the range value of 1 to 3 are more common in F2 (uniformity of cell size) among the benign class and the range value of 4 to 10 was more common among the malignant class. The statistical analysis of features normal nuclei, uniformity of cell size, bland chromatin, and bare nuclei are represented in Figures 2A, 2B, 2C, and 2D, respectively.

Figure 2 shows the histogram of frequency values and membership functions (MF) of each feature in the BC dataset in two classes (benign/malignant). Also, the output of membership functions is demonstrated in Figure 2. Membership functions for input data in a fuzzy system are determined based on the frequency of data in the selected features. More details of the input and output of the fuzzy system are presented in Table 2.

Table 2 shows the input features of the fuzzy system, range, linguistic variables, and type of their membership functions. It is worth noting that the de-fuzzifier method in this FIS is the centroid.

4.3. Sparse Representation for Classification (SRC)

The SRC is considered as an active research field for classification purposes. The signals potentially represent

a linear combination of atoms in an over-complete dictionary. Therefore, SRC can be applied to various classification tasks. It is based on the hypothesis that a prototype is represented in linear matter with a few tutorials of the same class. More technical description of SRC can be found in (30) and (31). Algorithm 2 (Box 2) shows how the SRC method in which the training data had two classes (malignant and benign $\psi = [\psi_1 \ \psi_2]$) and the testing data is Y . The relationship between the training and testing data is reported as $Y = \alpha_{i,1} \psi_{i,1} + \alpha_{i,2} \psi_{i,2} + \dots + \alpha_{i,n_i} \psi_{i,n_i}$, where i is the number of classes and n_i is the amount of training data in class i . To achieve the coefficient \hat{A} in the minimization problem is the main aim of the sparse algorithm in this study. As l_0 Minimization is NP-Hard problem, the l_1 Minimization is replaced with l_0 Minimization for solving this optimization problem.

There are various optimization methods for solving the minimization problem in the sparse representation algorithm. Some of them include Orthogonal Matching Pursuit (OMP), Orthogonal Recurrent Matching Pursuit (ORMP), Lasso, FOCUSS, Basic Pursuit, and Pinv (complete name) [19, 20]. More details about the parameters are reported in the next section.

5. Experimental Results

Here we use WBCD dataset for the evaluation of the proposed method. According to Section 4.2, the ICA feature selection method was applied for selecting the most effective feature. Therefore, four influencing features (F1: normal nuclei, F2: uniformity of cell size, F7: bland chromatin, F9: bare nuclei) were selected among the nine features with regularization parameters as mentioned in Table 3.

The value of 1 in Table 3 indicates the selected features, and the dash indicates the unselected features. It means that all of the features were selected in row 1 (without feature selection) of Table 3, and features 1, 2, 7 and 9 were selected based on the ICA feature selection method in row 2. As shown in Table 4, the accuracy rate of BC detection is evaluated as follows:

State 1: Without feature selection, all of the features (9 features) in the dataset were considered as usual (the weight of each feature is 1). In this state, the accuracy rate was 88.86%.

State 2: ICA feature selection: Four features were selected by the ICA feature selection method. The accuracy rate of 87.19% was achieved only with the four selected features (weight of the selected features was 1 while others was 0). In this state, the unselected features were eliminated, and just four features remained for diagnosis and classification.

Table 1. The Result of the ICA Method on BC Dataset

Regularization Parameters	Imperialist	Position	Cost	Number of Colonies
Population size = 200; revolution rate = 0.3; epoch = 50; number of Initial IMP = 20; number of IMP after 50 epochs = 16	1	[1,1,1,1,0,0,1,0]	0.6924	11
	2	[1,1,1,1,1,0,1,0]	0.6967	11
	3	[1,1,0,0,1,0,0,0]	0.6924	10
	4	[1,1,0,1,0,0,1,1]	0.6967	7
	5	[1,1,0,1,1,0,1,1]	0.6924	10
	6	[0,0,1,0,1,0,0,1]	0.6924	10
	7	[1,1,0,1,1,1,1,0]	0.6924	10
	8	[1,1,0,0,0,0,1,0,1]	0.6981	23
	9	[1,1,0,1,0,0,1,1,0]	0.6952	13
	10	[1,1,0,1,0,1,1,1,0]	0.6909	9
	11	[1,0,0,1,1,0,1,1,0]	0.6924	6
	12	[0,1,1,1,1,1,1,0,0]	0.6952	12
	13	[0,0,1,1,0,0,1,0,0]	0.6895	8
	14	[1,0,0,1,0,1,1,0,1]	0.6924	17
	15	[1,0,1,0,1,1,0,1,1]	0.6938	9
	16	[1,0,1,1,1,0,0,0,0]	0.6952	10
	17	[1,1,0,1,0,0,0,0,0]	0.6953	7

Table 2. Input/Output of the Fuzzy System in the Tuning of the Selected Attributes

Input/output	Input Linguistic Variables	Selected by ICA	Range	Parameter Name	Type of Membership Function
Input	(Low),(High)	Yes	1-10	Feature 1: Clump Thickness	gbelfm
	(Low),(Moderate),(High)	Yes	1-10	Feature 2: Uniformity of Cell Size	
	-	No	1-10	Feature 3: Uniformity of Cell Shape	
	-	No	1-10	Feature 4: Marginal Adhesion	
	-	No	1-10	Feature 5: Single Epithelial Cell Size	
	-	No	1-10	Feature 6: Bare Nuclei	
	(Low),(Moderate),(High)	Yes	1-10	Feature 7: Bland Chromatin	
	-	No	1-10	Feature 8: Normal Nucleoli	
	(Low),(High)	Yes	1-10	Feature 9: Mitoses	
Output	(V-Low),(Low),(Moderate),(High),(V-High)	-	1-10	Fuzzy Weight	

State 3: Manual weighting for the selected features: In this state, the accuracy rates were achieved through weighting the selected features (F1, F2, F7, and F9). This weighting was applied to distinguish the effect of the selected features. Weighting in this state was performed manually in the range between 1 and 2 (the weight of the selected feature was between 1 and 2 while others was 1).

State 4: Fuzzy weighting for selected features: The manual weighting in state 3 was not reasonable for each dataset

and needed the repetition and experience for selecting the best weight to achieve the best result. Therefore, the tuning of attributes was performed by a fuzzy system to weight the selected features. The fuzzy system produced a weight as an output of the fuzzy rule-based system. The accuracy of 95.45% was achieved when the fuzzy weight was applied to the selected features.

According to the result of step 4 in Table 4, γ is the fuzzy weight as an output of the fuzzy system. This weight was

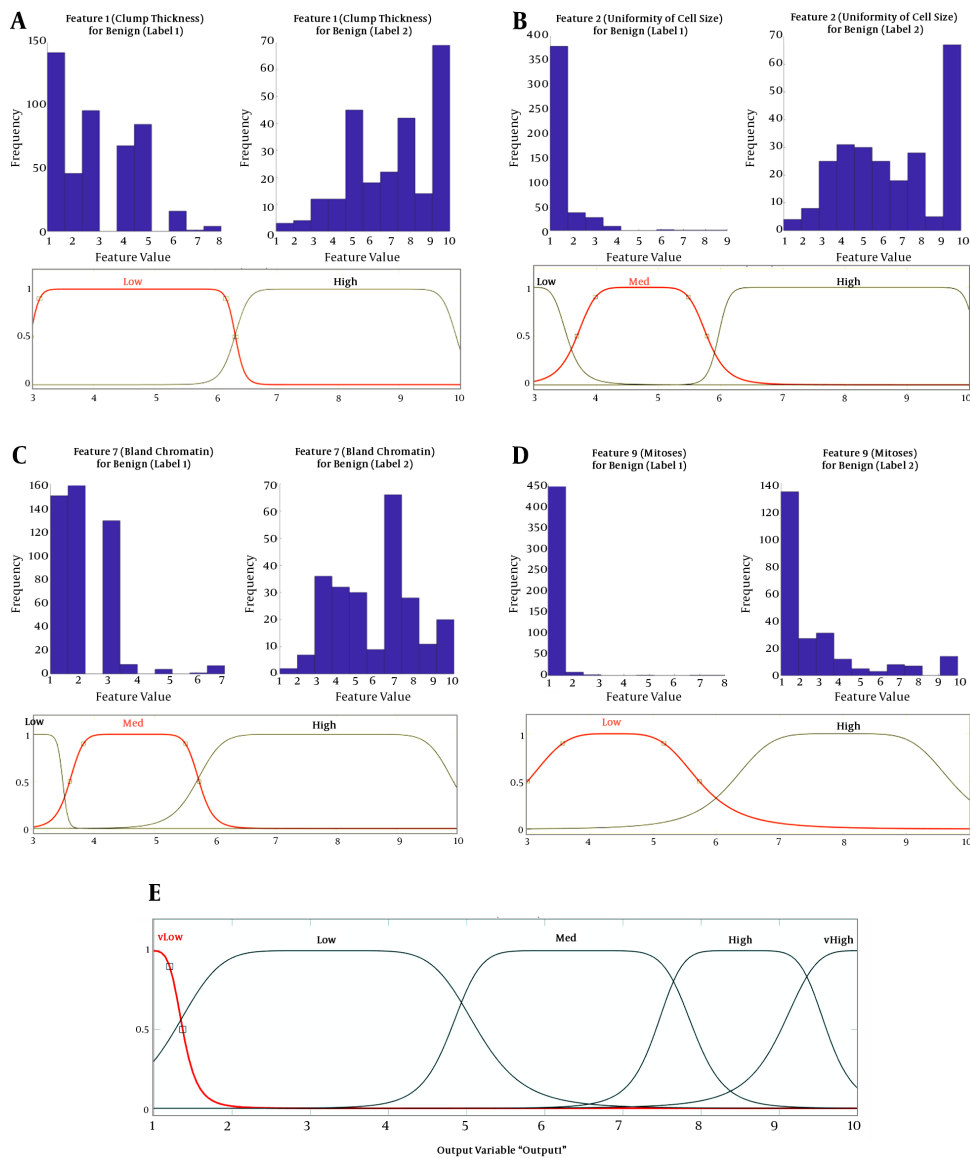


Figure 2. The histogram of frequency values and membership (MF) functions of each feature in the BC dataset in two classes (benign/malignant). A, The frequency values and MF of normal nuclei; B, The frequency values and MF of uniformity of cell size; C, The frequency values and MF of bland chromatin; D, The frequency values of bare nuclei; E, The membership function of the fuzzy output

calculated through inference fuzzy rule-based system between F1, F2, F7, and F9 features as the inputs of the fuzzy system. Several numbers of samples and their calculated fuzzy weights in the current dataset are shown in Table 5.

As shown in Table 5, the output of the fuzzy system was applied to the selected features, and other features remained unchanged. After feature selection and tuning the attributes using a fuzzy system, sparse representation was performed for each sample. Different sparse optimization methods such as OMP, Pinv, FOCUSS, Basic Pursuit, and

ORMP were evaluated in this study. The values of evolution methods including RMSE, R, R², accuracy, and true/false percentage on the detection of malignant/benign for several numbers of sparse optimization methods are summarized in Table 6.

The OMP (Orthogonal Matching Pursuit), sparse optimization method, has a desirable result compared to the other optimization methods in both the feature selection and weighted functions for the ICA feature selection. True positive, true negative, false positive, and false negative are

Box 2. Algorithm 2 - Sparse Representation for Classification (SRC)

Sparse Representation for Classification (SRC) Algorithm	
1. Input	Training data ($\psi = [\psi_1 \ \psi_2]$), $\psi_1 =$ training data for class malignant, $\psi_2 =$ training data for class benign; testing data (Y)
2. Problem definition	$Y = \alpha_{i,1} \psi_{i,1} + \alpha_{i,2} \psi_{i,2} + \dots + \alpha_{i,n_i} \psi_{i,n_i}, i = 1, 2$ (number of classes), $n_i =$ number of training data in class i ; $Y = A \psi, (Y = A \psi + z, z = \text{noise})$
3. Coefficient vector	$A = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0] \in \mathbb{R}^n$; $\hat{A} = [\hat{A}_1 \ \hat{A}_2], \hat{A}_i = \left\ \{\alpha_{i,j}\}_{j=1}^{n_i} \right\ _2$
4. l_0 Minimization	$\arg \min \left\ \hat{A} \right\ _0$ subjected to $\ Y - A \psi\ _2 < \varepsilon$; (l_0 Norm NP-Hard problem and convert to l_0 Norm)
5. l_1 Minimization	$\arg \min \left\ \hat{A} \right\ _0$ subjected to $\ Y - A \psi\ _2 < \varepsilon$; $\arg \min \left\ \hat{A} \right\ _1 + \ Y - A \psi\ _2, \lambda =$ regularization parameter
6. Calculate residual	$r_i(g) = \ Y - A \delta_i(\psi)\ _2, \quad i = 1, 2$
7. Output	Estimated labels for testing data; $\arg \min r_i(Y)$

Table 3. The Selected Feature via the ICA Feature Selection Algorithm and ICA Regularization Parameter

Method	Parameter	F1	F2	F3	F4	F5	F6	F7	F8	F9
Without feature selection	-	1	1	1	1	1	1	1	1	1
ICA feature selection	Population size = 200; revolution rate = 0.3; epoch = 50; number of IMP = 20	1	1	-	-	-	-	1	-	1

Table 4. The Procedure and Weighing of Features for Accuracy in Breast Cancer Detection

Method	Weight	F1	F2	F3	F4	F5	F6	F7	F8	F9	Accuracy	
State 1: Without feature selection	$\alpha = 1$	$\alpha \times F1$	$\alpha \times F2$	$\alpha \times F3$	$\alpha \times F4$	$\alpha \times F5$	$\alpha \times F6$	$\alpha \times F7$	$\alpha \times F8$	$\alpha \times F9$	88.86	
State 2: ICA feature selection	$\alpha = 0; \beta = 1$	$\beta \times F1$	$\beta \times F2$	$\alpha \times F3$	$\alpha \times F4$	$\alpha \times F5$	$\alpha \times F6$	$\beta \times F7$	$\alpha \times F8$	$\beta \times F9$	87.19	
State 3: Manual weighing for ICA selected features	$\alpha = 1$	$\beta = 1.6$	$\beta \times F1$	$\beta \times F2$	$\alpha \times F3$	$\alpha \times F4$	$\alpha \times F5$	$\alpha \times F6$	$\beta \times F7$	$\alpha \times F8$	$\beta \times F9$	89.17
		$\beta = 1.7$										89.38
		$\beta = 1.8$										89.79
		$\beta = 1.9$										89.9
		$\beta = 2$										90
State 4: Fuzzy weighing for ICA selected features	$\alpha = 1$	$\gamma =$ Fuzzy weight (Table 3)	$\gamma \times F1$	$\gamma \times F2$	$\alpha \times F3$	$\alpha \times F4$	$\alpha \times F5$	$\alpha \times F6$	$\gamma \times F7$	$\alpha \times F8$	$\gamma \times F9$	95.45

Table 5. Calculated Weights as the Output of a Fuzzy System for Several Numbers of Samples in the Dataset

Input Fuzzy System (a Selected Feature with ICA Feature Selection Algorithm)				Output (Calculated Weight with Fuzzy Rule)	$\gamma \times$ Features (Data for Classification)								
F1	F2	F6	F9	Weight (γ)	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	1	3	1	7.44	7.4	7.4	1	1	2	22.3	3	1	7.4
8	7	9	4	3.21	25.7	22.4	5	10	7	28.9	5	5	12.8
7	4	1	1	5.12	35.8	20.5	6	4	6	5.12	4	3	5.12
4	1	1	1	9.36	37.4	9.36	1	1	2	9.36	2	1	9.36
4	1	1	1	9.36	37.4	9.36	1	1	2	9.36	3	1	9.36
10	7	10	2	3.19	31.9	22.3	7	6	4	31.9	4	1	6.38
6	1	1	1	7.79	46.74	7.79	1	1	2	7.79	3	1	7.79
7	3	10	4	4.64	32.51	13.9	2	10	5	46.4	5	4	18.5
10	5	7	1	6.25	62.58	31.2	5	3	6	43.8	7	10	6.25

the true measurement percentage of detection of malignant, true percentage detection of the benign, false percentage of detection of the malignant and false percentage of detection of benign, respectively.

Figure 3 shows the accuracy rates among sparse optimization methods, both in sensitivity and in specificity (Equations 1 The best approach has the highest value of sen-

sitivity and specificity. The range of values for sensitivity and specificity are between 0 and 1.

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (2)$$

Table 6. The Results of the Proposed Method for Weighting the Selected Features and Without Feature Selection

Method	Sparse Optimization Method	True Positive (Truly Detect Malignant)	True Negative (True Detect Benignly)	False Positive (False Detect Malignant)	False Negative (False Detect Benign)	RMSE	R	R ²	Accuracy
Without feature selection	OMP	94.32	83.40	5.03	16.6	0.7157	0.1495	0.1910	88.86
	Pinv	53.49	99.17	46.51	0.83	0.7649	0.0871	0.2101	76.33
	FOCUSS	9.17	98.75	90.83	1.25	0.8032	0.0427	0.0572	53.96
	Basic Pursuit	8.73	98.75	91.27	1.25	0.8032	0.0398	0.0547	53.74
	ORMP	94.97	82.15	5.03	7.75	0.7116	0.1437	0.1882	88.56
Weighting the selected features	OMP	96.72	94.19	3.28	5.81	0.7285	0.1806	0.1940	95.45
	Pinv	56.33	95.85	43.67	4.15	0.7546	0.0718	0.2160	76.09
	FOCUSS	43.01	94.60	56.99	5.4	0.7640	0.0499	0.1926	68.80
	Basic Pursuit	26.63	93.36	73.37	6.64	0.7640	0.0499	0.1926	60
	ORMP	92.28	92.53	7.72	7.47	0.7246	0.1632	0.1926	94.40

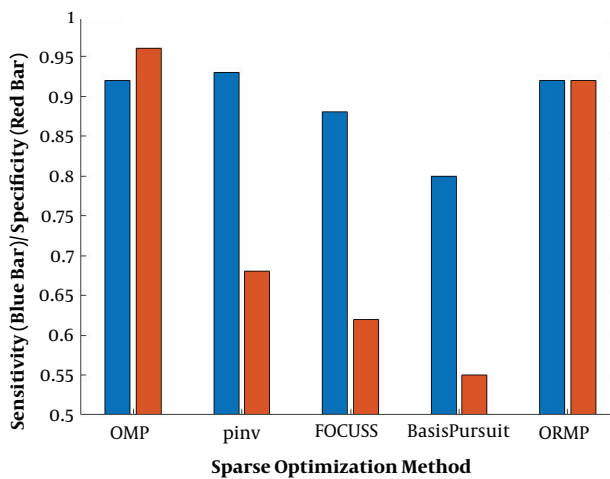


Figure 3. Result of calculated sensitivity and specificity value on several sparse optimization methods

As shown in Figure 3, OMP had the highest value in both sensitivity and specificity. Figure 4 shows ROC Curves for OMP, ORMP, FOCUSS, Pinv, and Basis Pursuit Sparse Optimization Methods on the classification of the selected features. Figure 4 shows the ROC plot, which measures the area under the ROC that shows how well a parameter can distinguish between malignant and benign classes.

According to Figure 4, the ROC curve for Orthogonal Matching Pursuit (OMP) Sparse Optimization on the selected features has the highest accuracy rate. After OMP, ROMP demonstrate the best result compared to the other methods. Figure 5 indicates the ROC Curves for OMP (the abbreviation is already introduced) Sparse Optimization with and without feature selection.

According to Figure 5, ROC Curves shows a better performance for OMP Sparse Optimization on the selected features with the fuzzy tuning compared to the other case, in which feature selection had not been performed. Table 7

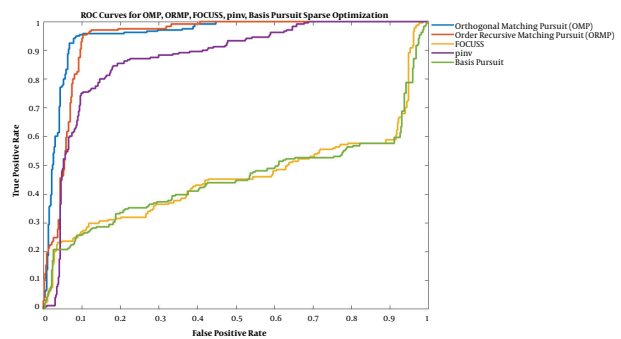


Figure 4. ROC Curves for OMP, ORMP, FOCUSS, Pinv, and Basis Pursuit Sparse Optimization Methods on the classification of the selected feature with ICA feature selection algorithm

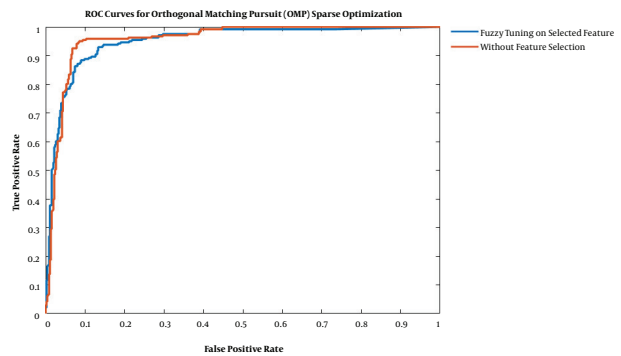


Figure 5. ROC Curves for OMP Sparse Optimization with and without feature selection

illustrates the performance of the proposed method compared with some of the primary techniques.

As shown in Table 7, our proposed method achieves 95.45 percentage of accuracy rate, which is the best compared to other primary techniques such as Multi-layer Perceptron (MLP) and Logistic Regression (LR). It also demonstrates better running time compared to MLP and LR.

Table 7. Performance of the Proposed Method Compared with Some of Basic Methods

Algorithm	Testing Accuracy	Running Time
MLP	91.2	1.94
Logistic regression	92.5	0.23
Proposed method	95.45	0.0034

6. Conclusions

This study introduced a new multi-phase feature selection method that uses a fuzzy system integrated with the ICA feature selection algorithm to robustly distinguish tumor samples from healthy samples. To choose the most influencing features, it employs the sparse algorithm to classify breast cancer. Our achieved results demonstrate the effectiveness of the proposed feature selection strategy. The experimental results were optimized by sparse representation method, and we found that the Orthogonal Matching Pursuit (OMP) was the best method in sparse representation with feature selection. The promising result of our proposed model suggests its possible usage as a substitute for medical invasive detection management. The proposed hybrid ICA-Fuzzy-SR method with 94.52% accuracy and 0.7285 of RMSE rate demonstrated excellent performance compared to current methods.

Acknowledgments

We would like to kindly thank Assistant Professor Abdollah (Iman) Dehzhangi (Department of Computer Science, Morgan State University, Baltimore, Maryland, US) for his very useful help in the revision of this manuscript. He put lots of time to review and revise the manuscript; however, because of the journal policy we were not able to have him as an author in the revised version of the paper.

Footnotes

Authors' Contribution: Shahab Shamshirband designed the study. Shahab Shamshirband, Irvan Masoudiasl, analyzed and visualized the results. Shahab Shamshirband, Shaghayeh Vahdat, and Somayeh Hessam wrote the paper. Shahab Shamshirband and Hamid Alinejad-Rokny revised the paper.

Conflict of Interests: There were no conflicts of interest.

Funding/Support: There is no grant number for this research. This paper has been supported by Ton Duc Thang University.

References

- Bhardwaj A, Tiwari A. Breast cancer diagnosis using Genetically Optimized Neural Network model. *Expert Syst Appl*. 2015;**42**(10):4611-20. doi:10.1016/j.eswa.2015.01.065.
- World Health Organization. *Breast cancer: Prevention and control*. 2012. Available from: <http://www.who.int/cancer/detection/breastcancer/en/>.
- Elouedi H, Meliani W, Elouedi Z, Ben Amor N, editors. A hybrid approach based on decision trees and clustering for breast cancer classification. *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of IEEE*. 2014. p. 226-31.
- Wolberg WH, Street WN, Mangasarian OL. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Anal Quant Cytol Histol*. 1995;**17**(2):77-87. [PubMed: 7612134].
- Schaefer G, Zaviscek M, Nakashima T. Thermography based breast cancer analysis using statistical features and fuzzy classification. *Pattern Recogn*. 2009;**42**(6):1133-7. doi: 10.1016/j.patcog.2008.08.007.
- Chen W, Giger ML, Bick U. A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol*. 2006;**13**(1):63-72. doi: 10.1016/j.acra.2005.08.035. [PubMed: 16399033].
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med*. 2001;**7**(6):673-9. doi: 10.1038/89044. [PubMed: 11385503]. [PubMed Central: PMC1282521].
- Yaghoobi H, Ghahramani Barandagh A, Mohammadi Z, editors. Breast cancer diagnosis using grey-level co-occurrence matrices, decision tree classification and evolutionary feature selection. *2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI): IEEE*. 2015. p. 317-24.
- Mahmoodabadi Z, Saniee Abadeh M. CADICA: Diagnosis of coronary artery disease using the imperialist competitive algorithm. *J Comput Sci Eng*. 2014;**8**(2):87-93. doi:10.5626/jcse.2014.8.2.87.
- Farokhian F, Yang C, Demirel H, Wu S, Beheshti I. Automatic parameters selection of Gabor filters with the imperialism competitive algorithm with application to retinal vessel segmentation. *Biocybernet Biomed Eng*. 2017;**37**(1):246-54. doi: 10.1016/j.bbe.2016.12.007.
- Paul AK, Shill PC, Rabin M, Murase K. Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. *Appl Intell*. 2017;**48**(7):1739-56. doi: 10.1007/s10489-017-1037-6.
- Yang J, Feng X, Angelini ED, Laine AF, editors. Texton and sparse representation based texture classification of lung parenchyma in CT images. *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2016. p. 1276-9.
- Yip CH, Taib NA. Breast health in developing countries. *Climacteric*. 2014;**17** Suppl 2:54-9. doi: 10.3109/13697137.2014.947255. [PubMed: 25131779].
- Kooi T, Litjens G, van Ginneken B, Gubern-Merida A, Sanchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;**35**:303-12. doi: 10.1016/j.media.2016.07.007. [PubMed: 27497072].
- Buyukavcu A, Albayrak YE, Goker N. A fuzzy information-based approach for breast cancer risk factors assessment. *Appl Soft Comput*. 2016;**38**:437-52. doi:10.1016/j.asoc.2015.09.026.
- Williams K, Adebayo Idowu P, Ademola Balogun J, Ishola Oluwaranti A. Breast cancer risk prediction using data mining classification techniques. *Trans Network Commun*. 2015;**3**(2). doi:10.14738/tnc.32.662.
- Diz J, Marreiros G, Freitas A. Applying data mining techniques to improve breast cancer diagnosis. *J Med Syst*. 2016;**40**(9):203. doi: 10.1007/s10916-016-0561-y. [PubMed: 27498205].
- Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst Appl*. 2014;**41**(4):1476-82. doi: 10.1016/j.eswa.2013.08.044.

19. Saybani MR, Wah TY, Aghabozorgi SR, Shamshirband S, Mat Kiah ML, Balas VE. Diagnosing breast cancer with an improved artificial immune recognition system. *Soft Comput.* 2016;**20**(10):4069-84. doi: [10.1007/s00500-015-1742-1](https://doi.org/10.1007/s00500-015-1742-1).
20. Saybani MR, Shamshirband S, Golzari S, Wah TY, Saeed A, Mat Kiah ML, et al. RAIRS2 a new expert system for diagnosing tuberculosis with real-world tournament selection mechanism inside artificial immune recognition system. *Med Biol Eng Comput.* 2016;**54**(2-3):385-99. doi: [10.1007/s11517-015-1323-6](https://doi.org/10.1007/s11517-015-1323-6). [PubMed: [26081904](https://pubmed.ncbi.nlm.nih.gov/26081904/)].
21. Sugiyama M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *J Mach Learn Res.* 2007;**8**(May):1027-61.
22. Martinez AM, Kak AC. PCA versus LDA. *IEEE Trans Pattern Anal Mach Intell.* 2001;**23**(2):228-33. doi: [10.1109/34.908974](https://doi.org/10.1109/34.908974).
23. Al-Shaikhli SDS, Yang MY, Rosenhahn B. Automatic 3D liver segmentation using sparse representation of global and local image information via level set formulation. *arXiv.* 2015:1508-21.
24. Li S, Jiang H, Yao YD, Yang B. Organ location determination and contour sparse representation for multiorgan segmentation. *IEEE J Biomed Health Inform.* 2018;**22**(3):852-61. doi: [10.1109/JBHI.2017.2705037](https://doi.org/10.1109/JBHI.2017.2705037). [PubMed: [28534802](https://pubmed.ncbi.nlm.nih.gov/28534802/)].
25. Jiang W, Yang X, Wu W, Liu K, Ahmad A, Sangaiah AK, et al. Medical images fusion by using weighted least squares filter and sparse representation. *Comput Electr Eng.* 2018;**67**:252-66. doi: [10.1016/j.compeleceng.2018.03.037](https://doi.org/10.1016/j.compeleceng.2018.03.037).
26. Frank A, Asuncion A. *UCI machine learning repository*. Irvine, CA: School of Information and Computer Science, University of California; 2010. Available from: <http://archive.ics.uci.edu/ml>.
27. Atashpaz-Gargari E, Lucas C, editors. Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. *IEEE Congress on Evolutionary Computation.* 2007. p. 4661-7.
28. Kalantari A, Kamsin A, Shamshirband S, Gani A, Alinejad-Rokny H, Chronopoulos AT. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing.* 2018;**276**:2-22. doi: [10.1016/j.neucom.2017.01.126](https://doi.org/10.1016/j.neucom.2017.01.126).
29. Miranda GH, Felipe JC. Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. *Comput Biol Med.* 2015;**64**:334-46. doi: [10.1016/j.compbiomed.2014.10.006](https://doi.org/10.1016/j.compbiomed.2014.10.006). [PubMed: [25453323](https://pubmed.ncbi.nlm.nih.gov/25453323/)].
30. Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. *Neurocomputing.* 2006;**70**(1-3):489-501. doi: [10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126).
31. Huang GB, Ding X, Zhou H. Optimization method based extreme learning machine for classification. *Neurocomputing.* 2010;**74**(1-3):155-63. doi: [10.1016/j.neucom.2010.02.019](https://doi.org/10.1016/j.neucom.2010.02.019).