Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2013) 21: 2079 – 2091

doi:10.3906/elk-1203-119

A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: RS + ELM

Yılmaz KAYA*
Department of Computer Engineering, Faculty of Engineering and Architecture, Siirt University, Siirt, Turkey

Received: 26.03.2012 • **Accepted:** 28.05.2012 • **Published Online:** 24.10.2013 • **Printed:** 18.11.2013

Abstract: Breast cancer is one of the leading causes of death among women all around the world. Therefore, true and early diagnosis of breast cancer is an important problem. The rough set (RS) and extreme learning machine (ELM) methods were used collectively in this study for the diagnosis of breast cancer. The unnecessary attributes were discarded from the dataset by means of the RS approach. The classification process by means of ELM was performed using the remaining attributes. The Wisconsin Breast Cancer dataset (WBCD), derived from the University of California Irvine machine learning database, was used for the purpose of testing the proposed hybrid model and the success rate of the RS + ELM model was determined as 100%. Moreover, the most appropriate attributes for the diagnosis of breast cancer were determined from the WBCD in this study. It is considered that the proposed method will be useful in similar medical practices.

Key words: Breast cancer, rough set, extreme learning machine, expert system, artificial intelligence

1. Introduction

The breast is an appendage of the skin covering the external part of our body and it includes lactating glands. Breast cancer (BC) is defined as the existence of cells progressing abnormally within the tissue of the breast that cannot be controlled. A group of cells growing or changing abnormally is called a tumor. Any tumor may be benign (not dangerous) or malignant (having the potential for being dangerous). BC is a cancer type that is very common in women and has a prevalence approximately 3 times higher than that of lung cancer [1]. The most common cancer types in women are BC and uterine cancer. In reports of the World Health Organization, it is anticipated that 1.2 million women will be diagnosed with BC every year [2,3]. BC has become an important problem for women. Therefore, uninterrupted research has been conducted for the purpose of the early diagnosis of BC. In medical science, various tests like clinical exams or mammography are performed for diagnosing BC. Mammography is X-ray photography of the breast taken using a low-dose X-ray machine. It is done for the purpose of determining the anomalies that are too small to detect by means of palpation. In addition to clinical tests, machine learning methods are also widely used for the early diagnosis of BC. Thanks to the success of these machine learning methods, they are widely utilized in the medical field by specialists. When the literature is examined, it can be seen that statistical and artificial intelligence techniques have been used successfully for the diagnosis of BC. The machine learning methods that were widely used for the diagnosis of BC and provided overachievement are discussed in the following section.

*Correspondence: yilmazkaya1977@gmail.com

A hybrid model based on rough set (RS) and extreme learning machine (ELM), the RS + ELM, is proposed in this study for diagnosing BC. The proposed model consists of 2 stages. The first is the stage of obtaining the optimal attribute subset representing the dataset. The second is the stage of performing the ELM classification process with the reduced attribute set so obtained. RS is a mathematical approach that was developed by Pawlak and is used for different purposes, like the selection of the attributes, implication of the attributes, reduction of the variables, implication of the decision-making rules, and pattern recognition [4]. The ELM is a single hidden layer feed-forwarded artificial neural network (ANN) model. In this model, the weights pertaining to the neurons available in the input layer and the threshold values pertaining to the neurons available in the hidden layer are produced randomly, while the outputs in the hidden layer are calculated analytically [5]. The most significant attribute of this model is that the learning process takes place very quickly.

The Wisconsin Breast Cancer datasets (WBCDs), derived from the University of California Irvine (UCI) machine learning database, were used for the purpose of testing the proposed hybrid model. The RS + ELM hybrid model was tried for different training-test percentages. When the results were examined, the highest success rate was determined as 100% when the training and the test datasets were selected at the percentage of 80% and 20%, respectively. As a result, we are of opinion that the proposed hybrid model will be a tool for assisting specialists in making decisions with respect to the patients at the final stage.

The content of this study is designed as follows. The other studies performed using the WBCD are summarized in the next section. The obtaining and the introduction of the dataset are explained in Section 3. The theoretical information with respect to the RS and ELM methods is provided in Section 4. The obtained experimental results are shared in Section 5. This study is discussed in the final section.

2. Studies for the diagnosis of BC

When the performed studies are examined, it is observed that the machine learning studies carried out using the WBCD are widespread and that high success rates were achieved in all of these performed studies.

Ster and Dobnikar [6] achieved a classification success rate of 96.80% using linear discriminant analysis. Pena-Reyes and Sipper [7] achieved a classification success rate of 97.36% in the study that they performed using a hybrid model based on fuzzy logic and the genetic algorithm (GA). The classification success rate achieved in the study by Setiono [8] was 98.10%. Abonyi and Szeifert [9] achieved a classification success rate of 95.57% using the controlled fuzzy set method. A classification success rate of 96.66% was achieved in the study by Kim et al. [10] using a fuzzy rule-based method. Sahan et al. [11] achieved a success rate of 99.14% using a hybrid model based on fuzzy artificial immunity and K-nearest neighbor in their studies. Polat and Güneş [12] achieved a success rate of 98.53% by least squares support vector machine (LS-SVM) in their studies. Karabatak and Ince [13] demonstrated a success rate of 97.4% by developing a model based on association rules and an ANN. Akay [2] reported a success rate of 99.51% using a model based on feature reduction and SVM. Kahramanli and Allahverdi [14] demonstrated a success rate of 92.31% using a model based on ANN and YBS. Marcano et al. [15] achieved a success rate of 99.26% in their studies. Hui et al. [16] reported a success rate of 99.41% using SVM and KK. As a result, high success rates were achieved in all of the studies performed for the diagnosis of BC by means of different machine learning methods.

3. Dataset

The WBCD, the dataset used in this study, was derived from the UCI machine learning database [17]. The dataset consists of 699 samples that were collected by Dr WH Wolberg at the University of Wisconsin-Madison

hospitals. A total of 16 instances were discarded from the dataset since they had missing observations and the RS + EML model was tested with the remaining 683 cases. The WBCD consists of 9 features and the values thereof range between 1 and 10. The target attribute was coded as benign (1 = benign) and malignant (0 = malignant). There are 444 benign cases and 239 malignant cases in the dataset. The attributes available in the dataset are detailed in Table 1.

Table 1. WBC data description of the attributes.

Feature	Code	Domain	Mean	Standard deviation
Clump thickness	A1	1–10	4.44	2.83
Uniformity of cell size	A2	1–10	3.15	3.07
Uniformity of cell shape	A3	1–10	3.22	2.99
Marginal adhesion	A4	1–10	2.83	2.86
Single epithelial cell size	A5	1–10	2.23	2.22
Bare nucleoli	A6	1–10	3.54	3.64
Bland chromatin	A7	1–10	3.45	2.45
Normal nucleoli	A8	1–10	2.87	3.05
Mitoses	A9	1–10	1.60	1.73

4. Method

4.1. Basic definitions of the RS theory

4.1.1. Information system

The information system in RSs is defined as $S = (U, Q, V)$. Here, $U = \{x_1, x_2, ...x_n\}$ indicates a finite nonempty universe. In this study, the universe is the set of patients. $Q = A \cup d$ indicates the finite nonempty attribute set and A indicates the set of case attributes pertaining to the patients. The set of case attributes covers the attributes given in Table 1, obtained from the patients, and is a vector of the attributes in the form of $A = \{a1,\ a2,\ldots an\}$. d is the decision attribute, indicating whether the patient has cancer or not. The information system is constituted by the combination of the case and the decision attributes. $V = \bigcup_{a \in A} V_a$ is the set of attributes pertaining to the a feature [17].

4.1.2. Indiscernibility relation

Observations cannot be discerned from each other due to the fact that a dataset is oversized or the obtained observations are similar to each other or identical. In such a case, the indiscernibility relation *IND(B)* for the attribute B can be written as follows, provided that $B \subseteq A$ [18]:

$$IND(B) = \{(x_1, x_2) \in U \times U : \forall a \in B, a(x_1) = a(x_2)\}. \quad (1)$$

Here, *IND(B)* is the B-indiscernibility relation. If x_1 and x_2 are included in the *IND(B)* set, the B attribute set, as well as x_1 and x_2, cannot be discerned from each other. The observation set (U = universe) can be divided into several equivalence classes in the form of U /*IND(B)*, according to the B-indiscernibility relation. These equivalence classes are shown in the form of $[x]_{IND(B)}$. All of the equivalence classes of *IND(B)* constitute the basic set of B. The equivalence classes, according to the decision-making attribute of the universe, form the value classes of the decision-making attribute.

4.1.3. Set approximations

The main objective in RSs is the creation of approximations using the *IND(B)* binary relation. It is the combination of the sets absolutely pertaining to X using the B-indiscernibility relation of X, provided that $X \subseteq U$, and it can be expressed as follows:

$$\underline{B}X = \cup\{x_i \in U | [x_i]_{IND(B)} \subseteq X\}. \tag{2}$$

Furthermore, the upper approximation can be written as follows [19,20]:

$$\bar{B}X = \cup\{x_i \in U | [x_i]_{IND(B)} \cap X \neq \varphi\}. \tag{3}$$

The upper and lower approximations pertaining to $X \subset U$ divides the universe (U) into 3 regions, namely the *POS(X)* positive region, *NEG(X)* negative region, and *BND(X)* bound region. The sets pertaining to these regions are calculated as follows [19]:

$$\begin{array}{l} POS(X) = \underline{B}X \\ NEG(X) = U - \overline{B}X \\ BND(X) = \overline{B}X - \underline{B}X \end{array} . \tag{4}$$

4.1.4. Attribute reduction and core attributes

Attribute reduction is the process of selecting the appropriate features from the attribute set for the purpose of explaining an information system with the minimum attributes. If *POS(B)* = *POS(A)*, provided that $B \subseteq A$, the information system can be explained with B consisting of the lower attribute number. Furthermore, an information system can have more than one reduced attribute set. The set obtained from the intersection of the reduced sets derived from an information system is called a core attribute set of the A attribute set [21,22]. The core attribute set can also be derived from the discernibility matrix.

4.1.5. Discernibility matrix

The discernibility matrix for the A case attributes in the S information system is $M(A) = (m_{ij})_{n \times n}$. $M(A)$ can be written as follows:

$$M(A) = \left\{ \begin{array}{l} \varphi \\ \{a \in A : a(x_i) \neq a(x_j)\} \end{array} \right. . \tag{5}$$

The $M(A)$ discernibility matrix has the feature of symmetry. Each component of $M(A)$ is constituted by the attribute set, making the x_i and x_j values different.

4.2. Extreme learning machine

The ELM developed by Huang et al. [5] will be described in this section. The ELM is a single hidden layer feed-forwarded ANN model, of which the input weights are calculated randomly, while the output weights are calculated analytically. Nondifferentiable or discrete activation functions can also be used in the hidden layer of the ELM, in addition to activation functions like sigmoidal, sine, Gaussian, and hard-limit [23]. Conventional feed-forwarded ANNs depend on some certain parameters like momentum or learning rate. The parameters, like the weights and threshold values in these types of networks, should be updated with gradient-based learning algorithms. However, for achieving a good performance, the learning process takes a long time and the error can be focused on a local point. Changing the momentum value may prevent the error from focusing on a local

point, but it will not have any influence on the long learning process. The input weights and threshold values are produced randomly, but the output weights are obtained analytically in ELM [23]. The ELM network is the customized form of a single hidden layer feed-forwarded ANN model. A single hidden layer feed-forwarded ANN is shown in Figure 1.

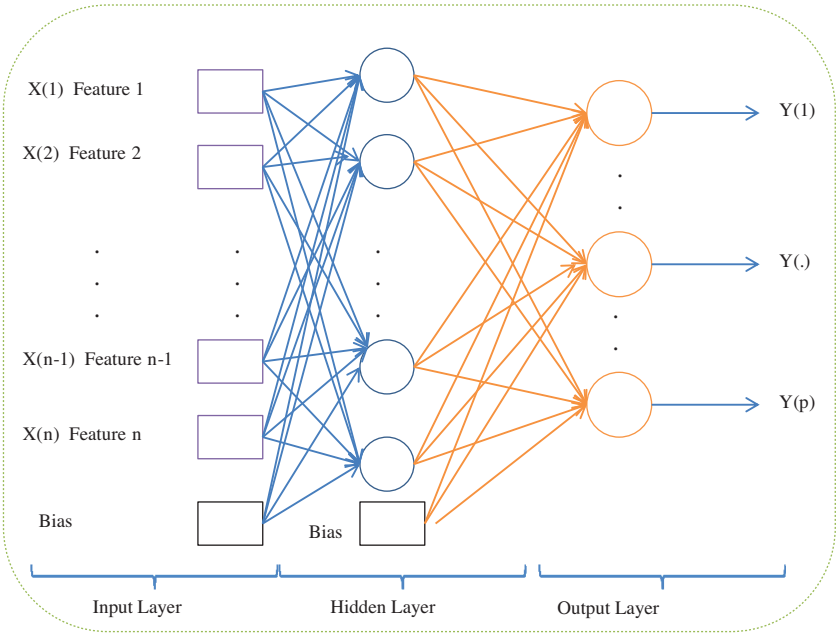

Figure 1. Feed-forwarded ANN.

Here, the mathematical statement of the network having M neurons in the hidden layer is expressed as follows, provided that $X = (X_1, X_2, X_3...X_N)$ indicates the input attributes and $Y = (Y_1, Y_2, Y_3.....Y_N)$ indicates the output attributes [24]:

$$\sum_{i=1}^{M} \beta_i g(W_i X_k + b_i) = O_k, k = 1, 2, 3...N. \tag{6}$$

Here, $W_i = (W_{i1}, W_{i2}, W_{i3}...W_{in})$ indicates the output weights in the input layer, $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3}.....\beta_{im})$ indicates the output weights in the hidden layer, b_i indicates the threshold values of the neurons in the hidden layer, and O_k indicates the output values of the network. $g(.)$ is the activation function [24]. The purpose in a network with N input is that the error is $\sum_{k=1}^{N} (o_k - Y_k) = 0$ or to obtain $\sum_{k=1}^{N} (o_k - Y_k)^2$ error. Therefore, Eq. (6) can be written as follows [5]:

$$\sum_{i=1}^{M} \beta_i g(W_i X_k + b_i) = Y_k, k = 1, 2, 3.....N. \tag{7}$$

With the above equation, it is possible to write the following [5]:

$$H\beta = Y. \tag{8}$$

Here, H, β, and Y can be expressed as follows [24]:

$$H(W_1,.....W_M;b_1,.....b_M;X_1,.....X_N) = \begin{bmatrix} g(W_1X_1+b_1) & ... & g(W_MX_M+b_M \\ . & & . \\ . & & . \\ g(W_1X_N+b_1) & ... & g(W_MX_N+b_M) \end{bmatrix} \tag{9}$$

and

$$\beta = \begin{bmatrix} \beta_1^T \\ . \\ . \\ \beta_M^T \end{bmatrix}_{M\times m} \text{ and } Y = \begin{bmatrix} Y_1^T \\ . \\ . \\ Y_N^T \end{bmatrix}_{N\times m} . \tag{10}$$

Here, H is the hidden layer output matrix. The training of the network in the conventional feed-forwarded ANN is to look for the LS solution in a linear equation $H\beta = Y$ in the ELM. The ELM algorithm can be summarized in 3 steps as follows briefly [5,25].

1. Step: The $W_i = (W_{i1}, W_{i2}, W_{i3}.....W_{in})$ input weights and hidden layer b_i threshold values are produced randomly.

2. Step: The H hidden layer output is calculated.

3. Step: The $\hat{\beta}$ output weights are calculated according to $\hat{\beta} = H^+Y$. Y is the target attribute.

4.3. Performance criteria

Accuracy, sensitivity, and specificity criteria were utilized in order to demonstrate the success of the RS + ELM method. Accuracy indicates the percentage of the accurately classified samples among all of the samples, sensitivity indicates the percentage of the accurately classified positive samples, and specificity indicates the percentage of the accurately classified negative samples [26]. These criteria are calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%, \tag{11}$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\%, \tag{12}$$

$$Specificity = \frac{TN}{FP+TN} \times 100\%. \tag{13}$$

Here, TP (true positive) as well as TN (true negative) indicate the true classifications and FP (false positive) as well as FN (false negative) indicate the false classifications.

TP: The number of samples that are not cancerous and are also indicated as healthy by the model.

TN: The number of samples that are cancerous and are also indicated as cancerous by the model.

FP: The number of samples that are cancerous, but are indicated as healthy by the model.

FN: The number of samples that are healthy, but are indicated as cancerous by the model.

4.4. Proposed hybrid model

A hybrid model based on the RS and ELM methods was used in this study to diagnose BC. The RS approach provides significant advantages in the determination of the relations between the attributes, reduction of the attributes, presentation of the importance of the attributes, and the establishment of the decision-making rules. In this study, the RS was used for the reduction of the attributes. At the classification stage, the ELM was used. Some important features related to the ELM can be listed as follows:

- ELM is quick.
- ELM has a generalizable performance.
- ELM does not need parameters like the learning rate or momentum, which are needed in conventional networks.
- ELM can use discrete or nondifferentiable activation functions in the hidden layer.

The diagram pertaining to the model used in this study is shown in Figure 2. The study consists of 7 blocks. The processes in these blocks can be summarized as briefly follows:

Block 1: Obtaining the BC data.

Block 2: Creation of the attribute sets, reduced by means of the RS from the obtained dataset.

Block 3: Determination of the optimal reduced attribute set.

Block 4: Dividing the dataset training-test partitions at different percentages, like 50%–50%, 70%–30%, and 80%–20%.

Block 5: Determination of the 50%–50%, 70%–30%, and 80%–20% partitions.

Block 6: Classification of the datasets by means of the ELM.

Block 7: Sharing the classification results.

5. Experimental results

5.1. Attribute reduction by means of the RS

If the number of case attributes in the datasets is high, it is frequently problematic to determine which attributes will be included in the model. It is not possible to make a decision about how many attributes or which attributes should be measured in the formation of the appropriate model by means of the conventional methods. The models formed with all of the measured attributes frequently lead to various problems. Therefore, the selection of the case attributes best explaining the decision-making attribute becomes crucial. The optimal attribute subsets obtained by means of the RS for the BC dataset discussed in this study are shown in Table 2. When the attribute subsets are examined, it is seen that the bare nucleoli (A6) attribute is available in all of the reduced attribute sets. The A6 attribute is the core attribute. There is a strong relation (high correlation) between the core attribute and the target attribute. It is seen that the attribute subsets consist of 4 or 5 attributes. The obtained attribute subsets will be used as the input for the ELM.

5.2. Form of the parameters for the ELM

The BC data were classified by means of the ELM using the attribute subsets obtained through the RS. The parameters pertaining to the ELM network used in this study are given in Table 3.

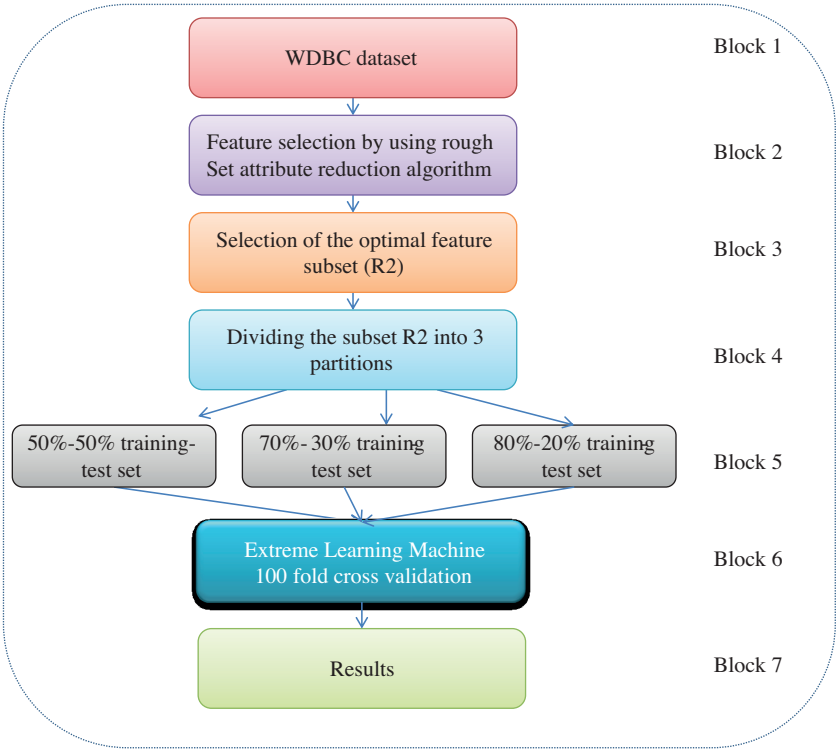

Figure 2. Diagram of the proposed hybrid model.

Table 2. Attribute subsets determined by the RS.

Reduct no.	Size	Features
R1	4	Clump thickness, marginal adhesion, bare nucleoli, bland chromatin
R2	4	Clump thickness, uniformity of cell shape, bare nucleoli, normal nucleoli
R3	4	Clump thickness, uniformity of cell shape, single epithelial cell size, bare nucleoli
R4	4	Clump thickness, uniformity of cell size, bare nucleoli, normal nucleoli
R5	5	Clump thickness, uniformity of cell size, single epithelial cell size, bare nucleoli, mitoses
R6	4	Uniformity of cell shape, marginal adhesion, bare nucleoli, normal nucleoli
R7	4	Clump thickness, single epithelial cell size, bare nucleoli, normal nucleoli
R8	5	Uniformity of cell size, single epithelial cell size, bare nucleoli, bland chromatin, mitoses
R9	5	Uniformity of cell size, single epithelial cell size, bland chromatin, normal nucleoli, bare nucleoli
R10	5	Uniformity of cell size, uniformity of cell shape, marginal adhesion, bare nucleoli, bland chromatin
R11	5	Uniformity of cell shape, marginal adhesion, bare nucleoli, bland chromatin, mitoses
R12	4	Single epithelial cell size, uniformity of cell shape, bare nucleoli, bland chromatin
R13	5	Single epithelial cell size, bare nucleoli, bland chromatin, normal nucleoli, mitoses

Table 3. Training parameters of the ELM network.

Number of layers	Input layer: 30 Hidden layer number: 1 Output layer: 1 Neuron number of hidden layers: 10...100
Activation functions	Tangent sigmoid, sigmoid, radial basis, triangular, sine
Learning algorithm	The ELM for single-hidden layer feedforward neural networks

The performance of the ELM network depends on the number of neurons in the hidden layer and the activation function to be used. Therefore, the suitability of the parameters in Table 3 was determined as a result of the trials. Activation functions like the sigmoid, tangent sigmoid, sine, and radial basis were used for the training and testing of the network to that end. The number of neurons in the hidden layer was determined by trials with increments within the range of 10–100. The optimal activation function and neuron number were decided according to the training and testing performance of the network. The optimal activation function in the classification of the BC was determined as tangent-sigmoid.

5.3. ELM classification results

Trials were conducted for the 50%–50%, 70%–30%, and 80%–20% training-test partitions, using all of the reduced attribute sets pertaining to BC. The success rates achieved for these training-test partitions are given in Table 4. The sensitivity, specificity, and accuracy rates obtained by means of the optimal reduced attribute set are given in Table 5. Figures 3, 4, and 5 illustrate the ELM performance values for the 50%–50%, 70%–30%, and 80%–20% training-test partitions, respectively. The classification performance values on the 50%–50%, 70%–30%, and 80%–20% training-test partitions were obtained after a 100-fold cross-validation.

Table 4. Performance values for the different reduced subset and training-test percentages.

Subset	50%–50% partition		70%–30% partition		80%–20% partition	
	Highest %	Average %	Highest %	Average	Highest %	Average %
R1	97.71	94.65	98.57	95.27	100.00	95.60
R2	97.71	95.52	99.52	96.15	100.00	96.19
R3	97.99	94.88	99.05	95.77	99.29	96.08
R4	98.57	95.58	99.05	96.34	99.29	96.63
R5	97.42	95.14	98.10	95.48	99.29	95.73
R6	97.13	94.54	98.10	95.29	99.29	95.31
R7	97.99	95.49	98.57	95.94	99.29	96.07
R8	96.56	94.01	97.62	94.64	98.57	94.91
R9	97.42	94.40	98.10	94.88	100.00	95.09
R10	97.13	94.58	98.10	95.28	99.29	95.57
R11	97.13	95.46	99.05	94.99	99.29	95.46
R12	96.85	94.17	97.62	95.10	99.29	95.41
R13	97.42	93.97	97.62	94.61	97.86	94.99

Table 5. Sensitivity, specificity, and accuracy rates for the R2 reduced subset.

Partition	Sensitivity %	Specificity %	Accuracy %
50%–50%	98.72	95.65	97.71
70%–30%	100.00	98.57	99.52
80%–20%	100.00	100.00	100.00

5.4. Attribute reduction by other methods

In this section, the significant features are obtained by principal component analysis (PCA), GA, linear discriminant analysis (LDA), stepwise forward and backward, a filter method relief, and chi-squared ranking methods, after the important features selected in the classification process are addressed through the ELM with the generated feature sets. The attributes obtained with the different reduction methods and classification results by the ELM are shown in Table 6.

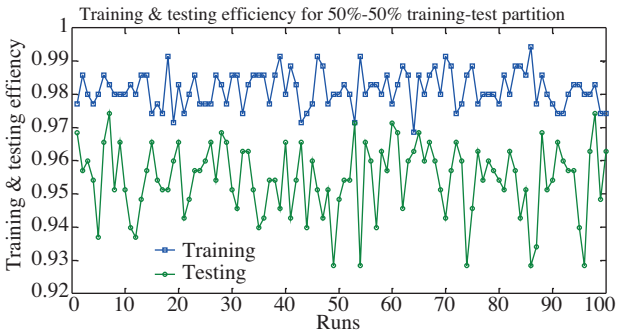

Figure 3. Training and test efficiencies for the 50%–50% training-test partition.

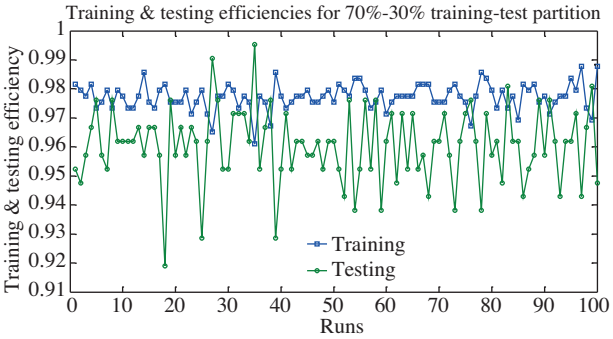

Figure 4. Training and test efficiencies for the 70%–30% training-test partition.

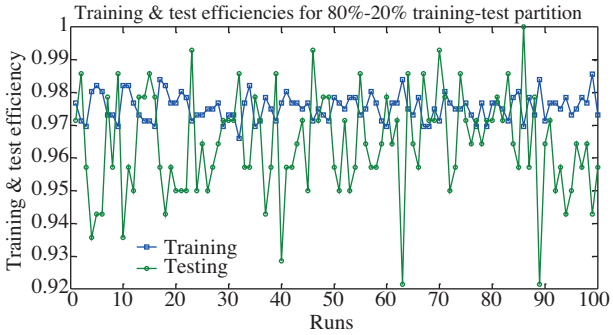

Figure 5. Training and test efficiencies for the 80%–20% training-test partition.

Table 6. Classification achievements for the different attribute reduction methods.

Attribute reduction method	Features	Accuracy % (max)	Accuracy % (average)
GA	A1, A2, A5, A6, A9	99.29	95.93
PCA	A1–A7	98.57	96.08
LDA	A1–A9	99.29	95.96
Stepwise forward	A1–A8	98.57	95.91
Stepwise backward	A1–A4, A6–A8	99.29	95.94
Relief	A1–A4, A6–A8	99.29	95.96
Chi-squared ranking	A1–A8	98.57	96.04
RS (reduction 2)	A1, A3, A6, A8	100.00	96.19

When Table 6 is examined, the best classification performance achieved with reduction 2 was obtained through the RS.

5.5. Comparison of the RS + ELM results with other studies in the literature

When the literature was examined, there were many machine learning methods formed using the WBCD. The studies conducted on the WBCD are given in Table 7, where it is seen that the RS + ELM achieved a considerable success in comparison with the other methods.

Table 7. Classification achievements of the previous studies for the diagnosis of BC.

Author	Method	Accuracy %
Quinlan [27]	C4.5	94.74
Hamiton et al. [28]	RAIC	95.00
Nauck and Kruse [29]	NEFCLASS	95.06
Abonyi and Szeifert [9]	SFC	95.57
Ster and Dobnikar [6]	LDA	96.80
Goodman et al. [30]	AIRS	97.20
Pena-Reyes and Sipper [7]	Fuzzy-GA1	97.36
Karabatak and Ince [13]	AR + NN	97.40
Abbas [31]	EANN	98.10
Setiono [8]	Neuro-rule	98.10
Polat and Güneş [12]	LS-SVM	98.53
Marcano et al. [15]	AMMLP	99.26
Hui et al. [16]	SVM + KK	99.41
Akay [2]	SVM-CFS	99.51
Present study(80%–20% training-test)	RS + ELM	100.00

6. Conclusion

Classification is an important tool used for diagnosing diseases in clinical practices. A support system related to medical decision-making was proposed in this study, using the RS and ELM models collectively for the diagnosis of BC. The RS was used for the reduction of the attributes, while the classification was made by means of the ELM, using the reduced attribute sets. For the performance test of the proposed RS + ELM method, the WBCD, which has been utilized widely by other researchers through different machine learning methods, was used. The dataset was divided into 50%–50%, 70%–30%, and 80%–20% training-test partitions and different practices were made for each partition during the study. The highest success rate was determined as 100% in the 80%–20% training-test partition. It was observed that the selection of the optimal attributes by means of the RS prior to the classification of the BC data positively influenced the success of the classification. As a result, we are of the opinion that the proposed model will be a tool for assisting specialists in making decisions at the final stage, as well as for different types of cancer.

References

[1] Wikipedia, Breast Cancer, available at https://en.wikipedia.org/wiki/Breast_cancer.

[2] M.F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis", Expert Systems with Applications, Vol. 36, pp. 3240–3247, 2009.

[3] Cancer. Available at. http://www.kanser.com.tr.

[4] C. Yumin, M. Duoqian, W. Ruizhi, "A rough set approach to feature selection based on ant colony optimization", Pattern Recognition Letters, Vol. 31, pp. 226–233, 2010.

[5] G.B. Huang, Q.Y. Zhu, C.K. Siew, "Extreme learning machine: theory and applications", Neurocomputing, Vol. 70, pp. 489–501, 2006.

[6] B. Ster, A. Dobnikar, "Neural networks in medical diagnosis: comparison with other methods", Proceedings of the International Conference on Engineering Applications of Neural Networks, pp. 427– 430, 1996.

[7] C.A. Pena-Reyes, M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis", Artificial Intelligence in Medicine, Vol. 17, pp. 131–155, 1999.

[8] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis", Artificial Intelligence in Medicine, Vol. 18, pp. 205–217, 2000.

[9] J. Abonyi, F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers", Pattern Recognition Letters, Vol. 14, pp. 2195–2207, 2003.

[10] M.W. Kim, J.W. Ryu, "Optimized fuzzy classification using genetic algorithm", Lecture Notes in Artificial Intelligence, Vol. 36, pp. 392–401, 2005.

[11] S. Şahan, K. Polat, H. Kodaz, S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis", Computers in Biology and Medicine, Vol. 37, pp. 415–423, 2007.

[12] K. Polat, S. Güneş, "Breast cancer diagnosis using least square support vector machine", Digital Signal Processing, Vol. 17, pp. 694–701, 2007.

[13] M. Karabatak, M. Ince, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications, Vol. 36, pp. 3465–3469, 2009.

[14] K. Humar, A. Novruz, "Rule extraction from trained adaptive neural networks using artificial immune systems", Expert Systems with Applications, Vol. 36, pp. 1513–1522, 2009.

[15] C.A. Marcano, D.J. Quintanilla, D. Andina, "WBCD breast cancer database classification applying artificial metaplasticity neural network", Expert Systems with Applications, Vol. 38, pp. 9573–9579, 2011.

[16] H.L. Chen, B. Yang, J. Liu, D.Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis", Expert Systems with Applications, Vol. 38, pp. 9014–9022, 2011.

[17] UCI Repository of Machine Learning Databases. Available at http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/.

[18] Y. Wang, M. Ding, C. Zhou, Y. Hu, "Interactive relevance feedback mechanism for image retrieval using rough set", Knowledge-Based Systems, Vol. 19, pp. 696–703, 2006.

[19] T. Wakaki, H. Itakura, M. Tamura, H. Motoda, T. Washio, "A study on rough set-aided feature selection for automatic web-page classification", Web Intelligence and Agent Systems: An International Journal, Vol. 4, pp. 431– 441, 2006.

[20] E. Stokić, V. Brtka, B. Srdić, "The synthesis of the rough set model for the better applicability of sagittal abdominal diameter in identifying high risk patients", Computers in Biology and Medicine, Vol. 40, pp. 786–790, 2010.

[21] C. Bae, W.C. Yeh, Y.Y. Chung, S.L. Liu, "Feature selection with intelligent dynamic swarm and rough set", Expert Systems with Applications, Vol. 37, pp. 7026–7032, 2010.

[22] J. Wang, K. Guo, S. Wang, "Rough set and Tabu search based feature selection for credit scoring", Procedia Computer Science, Vol. 1, pp. 2425–2432, 2010.

[23] S. Suresh, S. Saraswathi, N. Sundararajan, "Performance enhancement of extreme learning machine for multi-category sparse data classification problems", Engineering Applications of Artificial Intelligence, Vol. 23, pp. 1149–1157, 2010.

[24] H.J. Rong, Y.S. Ong, A.H. Tan, Z. Zhu, "A fast pruned-extreme learning machine for classification problem", Neurocomputing, Vol. 72, pp. 359–366, 2008.

[25] N.Y. Liang, P. Saratchandran, G.P. Huang, N. Sundararajan, "Classification of mental tasks from EEG signals using extreme learning machine", International Journal of Neural Systems, Vol. 16, pp. 29–38, 2006.

[26] L.Á. Menéndez, F.J. Cos Juez, L.F. Sánchez, R.J.A. Álvarez, "Artificial neural networks applied to cancer detection in a breast screening programme", Mathematical and Computer Modelling, Vol. 52, pp. 983–991, 2010.

[27] J.R. Quinlan, "Improved use of continuous attributes in C4.5", Journal of Artificial Intelligence Research, Vol. 4, pp. 77–90, 1996.

[28] H.J. Hamilton, N. Shan, N. Cercone, "RIAC: a rule induction algorithm based on approximate classification", International Conference on Engineering Applications of Neural Networks, University of Regina, 1996.

[29] D. Nauck, R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data", Artificial Intelligence in Medicine, Vol. 16, pp. 149–169, 1999.

[30] D.E. Goodman, L. Boggess, A. Watkins, "Artificial immune system classification of multiple-class problems", Proceedings of the Artificial Neural Networks in Engineering, pp. 179–183, 2002.

[31] H. Abbas, "An evolution artificial neural networks approach for breast cancer diagnosis", Artificial Intelligence in Medicine, Vol. 25, pp. 265–281, 2002.