

A novel and reliable computational intelligence system for breast cancer detection

Amin Zadeh Shirazi¹ · Seyyed Javad Seyyed Mahdavi Chabok¹ · Zahra Mohammadi²

Received: 14 March 2017 / Accepted: 28 August 2017 / Published online: 11 September 2017
© International Federation for Medical and Biological Engineering 2017

Abstract Cancer is the second important morbidity and mortality factor among women and the most incident type is breast cancer. This paper suggests a hybrid computational intelligence model based on unsupervised and supervised learning techniques, i.e., self-organizing map (SOM) and complex-valued neural network (CVNN), for reliable detection of breast cancer. The dataset used in this paper consists of 822 patients with five features (patient's breast mass shape, margin, density, patient's age, and Breast Imaging Reporting and Data System assessment). The proposed model was used for the first time and can be categorized in two stages. In the first stage, considering the input features, SOM technique was used to cluster the patients with the most similarity. Then, in the second stage, for each cluster, the patient's features were applied to complex-valued neural network and dealt with to classify breast cancer severity (benign or malign). The obtained results corresponding to each patient were compared to the medical diagnosis results using receiver operating characteristic analyses and confusion matrix. In the testing phase, health and disease detection ratios were 94 and 95%, respectively. Accordingly, the superiority of the proposed model was proved and can be used for reliable and robust detection of breast cancer.

Keywords Medical computing · Breast cancer · Medical diagnosis · Computational intelligence · Complex neural network · SOM

✉ Amin Zadeh Shirazi
amin.zadeshirazy@gmail.com; a.zadehshirazi@mshdiau.ac.ir

¹ Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Mashhad, Iran

² Department of Artificial Intelligence, Imam Reza International University, Mashhad, Iran

1 Introduction

Cancer is one of the main reasons of the morbidity and mortality worldwide, and in 2012, 14 million new cancerous patients were diagnosed and 8.2 million patients died of cancer [1]. In 2012, 1.7 million women with breast cancer were diagnosed, while there are 6.3 million women diagnosed to have cancer in 5 years before 2012 [2]. In 2012, having breast cancer was 20% and morbidity and mortality increased by 14% relative to 2008 [2].

It is completely obvious that breast cancer grows fast and extensively particularly in women. It is important to propose breast cancer prediction methods for early diagnosis and treatment in order to minimize the risk of metastasis of the cancerous tissues. In medical areas, the artificial neural network (ANN) has attracted considerable attentions to it. Considering high capability, this approach is used in a range of decision making. Saritas [3, 4], Saad [5], Groshev [6], and Wahab [7] used the ANN in cancer diagnosis. Many researchers used the ANN along with other techniques for breast cancer diagnosis. Karabatak et al. [8] used the integrative ANN and expert system approach based on association rules. They used expert system based on the association rules in order to reduce the breast cancer database and applied ANN for smart classification, and the results obtained from their proposed approach was compared to artificial neural network. Huang et al. [9] diagnosed the breast cancer using ANN classification with entropy based on feature selection. For ANN training, they reduced a set of data using feature selection and used Levenberg-Marquardt's model for training the ANN and employed the particle swarm optimization (PSO) for determination of the optimal ANN parameters. Senapati et al. [10] investigated the wavelet linear local neural network for breast cancer diagnosis such that the wavelet linear local neural network was used for recognition of the breast cancer, and they

optimized this approach by training its parameters using performance recursive least squares.

In addition to ANN, many researchers used another computing instrument for prediction and classification of the breast cancer patterns. Using machine learning models in genetic programming, Ludwig and Roos [11] predicted the breast cancer. They investigated the linear programming approach and genetic programming. Chao et al. [12] also used the SVM and logistic regression and decision tree for prediction of breast cancer survival. They employed a model for level of survival in patients with breast cancer using 10-fold validation for state recognition. Shin et al. [13] investigated the prediction of low-risk breast cancer tumors by perfusion parameters and apparent diffusion coefficient. They developed an empirical model for prediction of low-risk tumor using analysis of the logistic regression and analysis system performance characteristic curve. Korkmaz et al. [14] used the texture of images of optical microscope and mammography using relative entropy via estimation of the kernel for breast cancer diagnosis. For early breast cancer diagnosis by mammography and histopathology images and three triangles, Jensen Shannon and Hellinger's approaches, they evaluated the results. Nugroho et al. [15] investigated machine learning-based methods to improve the detection of breast cancer using mammography. They proposed a model based on naive Bayes and sequential minimal optimization (NB-SMO) to detect breast cancer.

It should be extremely noted that breast cancer is a multi-step and complicated disease with particular biological features and clinical behaviors [16] due to the abnormal proliferation of breast cells. Certainly, accurately diagnosing the breast cancer leads to minimization of morbidity and mortality and survival rate increase. Based on the fact that precision of the disease diagnosis is low and time consuming in traditional approach, it is necessary to have an efficient, accurate, and fast for helping the physicians in diagnosis and prediction of this disorder.

The innovation of this paper is dealing with the above-mentioned issues using a hybrid computational model based on unsupervised learning technique (self-organizing map) and supervised learning (complex-valued neural network) resulting in a reliable detection of breast cancer. Self-organizing map (SOM) and complex-valued neural network (CVNN) are the two popular and practical methods in the both fields of industry and medicine for clustering and classification purposes, respectively [17–21].

The rest of this paper is organized as follows. In Section 2, the methods, the dataset used along with the five most important features which are significant in the detection of breast cancer, will be explained. Section 3 describes the obtained results in detail. Then, Section 4 examines the proposed model (SOM-CVNN) and compares the obtained results with those of other models. Finally, in Section 5, conclusion and future work are described.

2 Methods

2.1 Data Acquisition and Feature Selection

In this article, the dataset used includes 822 patients, which was provided by M. Elter, R. Schulz-Wendtland, and T. Wittenberg [22]. There are novel techniques such as ultrasonography and MRI. However, due to their lack of capability in screening, they cannot be replaced with the traditional mammography. Mammography can diagnose the smallest abnormalities in breast tissue even in 0.5 cm, whereas breast mass in 1-cm size is not diagnosed by physicians. A mass can be either benign or malignant. Benign masses do not pose a threat health and the cells do not multiply rapidly. However, some types of benign breast conditions are linked to breast cancer risk. For example, proliferative lesions with atypia are associated with the greatest breast cancer risk [23]. Malignant masses have the potential to be dangerous for patients. They grow quickly and invade the nearby tissues, organs, and other parts of the body which can cause damage. As can be seen in Fig. 1, a mass is a space seen in two different projections. If a potential mass is seen in only a single projection, it should be called an “asymmetry” until its three-dimensionality is confirmed. A number of mammographic mass features are the shape of breast mass, density of breast mass, and margin of breast mass. The shape of a mass is either round, oval, or irregular. The margin of a lesion can be circumscribed (historically well-defined and is known as a benign mass), obscured or partially obscured, microlobulated, indistinct (historically ill-defined), and speculated. One of the most significant features that specify a malignant mass cancer is an irregular or speculated margin [24]. The density of a mass is related to the expected attenuation of an equal volume of fibro glandular tissue which has been proven to be a risk factor in breast cancer [25]. High density is associated with malignancy. It is extremely rare for a breast cancer tissue to be of low density.

On the other hand, American Radiology College registered the Breast Imaging Reporting and Data System (BI-RADS) as the diagnosis, prevention, and treatment guideline for the patients. BI-RADS is an instrument for qualitative expression and level of the risk in breast mammography and can reduce confusion in breast imaging interpretations. Employing this system makes the mammogram report reading for non-radiologists a standard and understandable issue and causes the improvement in relationship between physician and radiologist [26]. BI-RADS was updated in 2011 and divided into five categories [27]. All of these features have been depicted in Fig. 1.

2.2 Procedural information

2.2.1 SOM

Kohonen developed SOM in 1980s [28]. SOM is a very popular artificial neural network model, which is based on the

unsupervised learning paradigm [29, 30]. It consists of three phases, namely, competition, cooperation, and adaptation. This technique is a tool that maps large-scale data to data with fewer dimensions by putting together the most analogous data in the form of distinct clusters. Clustering and data compression are original applications of SOM. SOM generates a mapping from a continuous high-dimensional input space Ψ onto a discretized low-dimensional output space λ by learning from past examples. The discrete output space λ comprises k neurons arranged as per a fixed topology (a 2D hexagonal or rectangular lattice). The weight vectors $W = (w_1, w_2, \dots, w_k)$ define the mapping $c(x) : \Psi \rightarrow \lambda$, and it assigns to a neuron index (known as competition phase) an input vector $x(t)$:

$$i^*(t) = \arg \min_i \|x(t) - W_i(t)\| \quad (1)$$

where $\|\cdot\|$ refers to the Euclidean distance and t is the discrete current iteration. It is noteworthy that the dimensionality of weight vectors and input patterns is the same.

To train the weight vectors, a competitive-cooperative learning rule is adopted. When an input vector is fed to the network, the weight vector of the winner neuron as well as that

of its neighbors are updated as follows (known as adaptation phase):

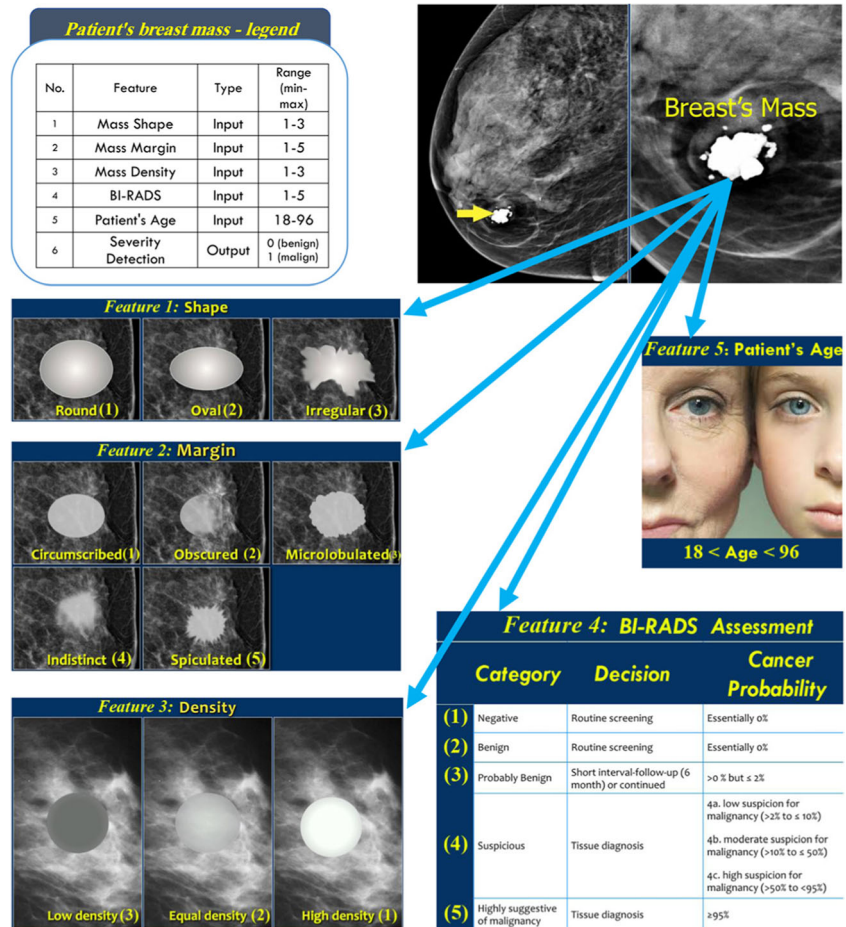
$$W_i(t + 1) = W_i(t) + \eta(t)h(i^*, i; t) \left[(x(t) - W_i(t)) \right] \quad (2)$$

Hence, the weight vectors of the adapted neurons slightly move toward the input vector. The learning rate η , which decreases with time exponentially, controls the magnitude of the movement. A neighborhood function h determines the number of neurons, which are impacted by this adaptation. The neighborhood function is typically unimodal, symmetric, and monotonically decreasing with distance to the winner increasing. The Gaussian function is a popular option (known as cooperation phase):

$$h(i^*, i; t) = \exp\left(-\frac{\|r_i(t) - r_{i^*}(t)\|^2}{2\sigma^2(t)}\right) \quad (3)$$

where $\|r_i(t) - r_{i^*}(t)\|$ is the distance between neurons i and i^* in the discrete output space λ , and $\sigma(t)$ is the radius of the neighborhood function at time t , which decreases exponentially to guarantee the decrease in the neighborhood size during training.

Fig. 1 The five most significant patient's features



The neighborhood function is chosen to embrace a large part of the output space λ at the onset of learning, and it is decreased little by little such that the winner is merely adapted toward the end of the process. When the global ordering of the weight vectors reaches a steady state, the map is said to have converged. The preservation of neighborhood relations, i.e., the mapping of nearby data vectors in the input space onto adjacent neurons in the output space, is a significant characteristic of the resultant map. The low-dimension output space can display the structure hidden in high-dimensional data, e.g., clusters and spatial relationships, because of the topology-preserving property.

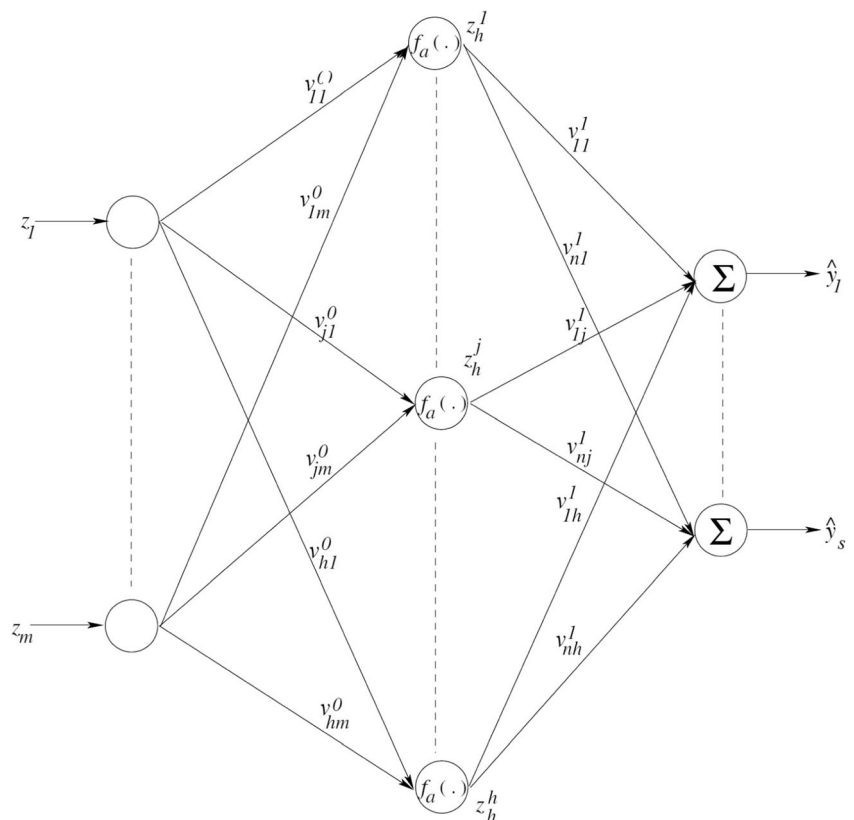
2.2.2 Min-max normalization technique

Sometimes, the values of input and output parameters are extremely low or high, so that raw data may not be suitable for use. Thus, raw data needs to undergo preprocessing. Therefore, scaling of data should be performed. One approach for scaling of data is performed with the following formula equation which normalizes the data to values between 0 and 1 [31, 32]:

$$X'_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

where X_i is the original value of parameter, X'_i is the normalized value of X_i , and X_{\min} and X_{\max} are the minimum and maximum values.

Fig. 2 Structure of a CVNN [33]



2.2.3 Transferring problem space

The phase transformation is used in order to transfer the input feature space with the real values to that with complex values as follows [20]:

If we have $x_j \in [a, b]$, so that $a, b \in R$, therefore, ϕ is defined as follows:

$$\phi = \frac{\Pi(x_j - a)}{b - a} \tag{5}$$

Then, supposing the constant magnitude and using the Euler equation, the complex variable z_j corresponding to the input variable x is as follows:

$$f_{R \rightarrow C} = z_j = e^{i\phi} = \cos(\phi) + (i)\sin(\phi) \tag{6}$$

As can be observed from the above equation, using this conversion, the input space of the features is mapped from $x \in [a, b]$ to $\phi \in [0, \frac{\Pi}{2}]$.

2.2.4 CVNN

These networks have been highly regarded since the 1980s, and it has been shown that CVNN makes it possible to solve complicated problems that cannot be solved by real-valued neural networks [19]. The networks developed in both

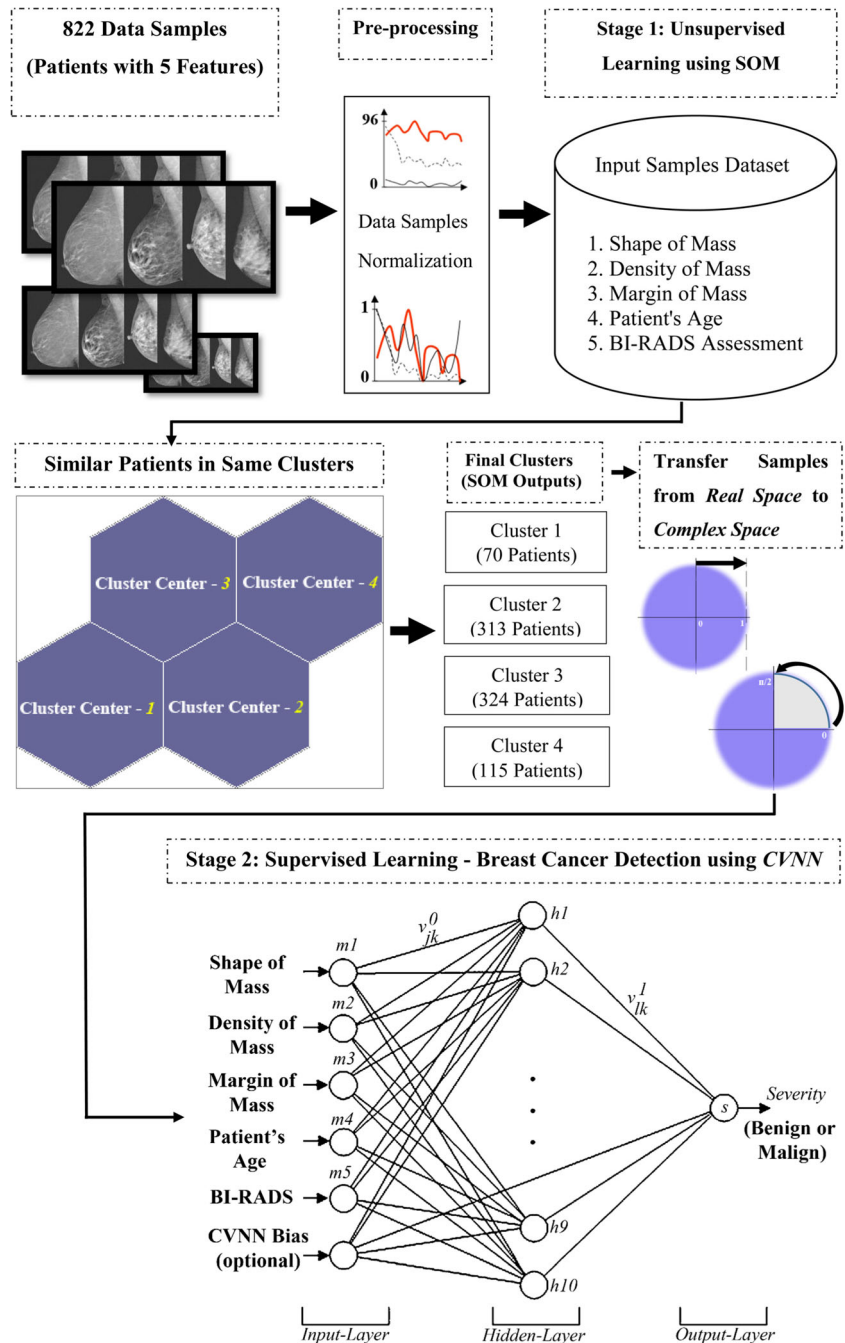
theoretical and practical fields so that, today, they have many applications in areas such as digital image processing and data processing [32]. Considering the set $\{(z_1, y_1), \dots, (z_t, y_t), \dots, (z_N, y_N)\}$, z_t and y_t are m - and s -dimensional inputs and outputs with complex values, respectively ($z_t \in C^m, y_t \in C^s$), that are applied to a CVNN. This network operates on a complex input with m dimensions $z_t = [z_1 z_2 \dots z_m]$, complex activation function $f_a(\cdot)$, complex input weight matrix V^0 , and complex output weight matrix V^1 to produce a complex output with s dimensions $\hat{y} = [\hat{y}_1 \hat{y}_2 \dots \hat{y}_s]$. The structure of CVNN is shown in Fig. 2. CVNN has m neurons in the input layer, h

neurons in the hidden layer, and s neurons in the output layer represented as $\xi^{m:h:s}$. Activation functions in the hidden and output layers also have complex values. The output of the j^{th} neuron is calculated in the hidden layer as follows:

$$Z_h^k = f_a \left(\sum_{j=1}^m V_{kj}^0 z_j \right); k = 1, 2, \dots, h \tag{7}$$

In the above equation, V_{kj}^0 is a complex weight that connects the j^{th} input neuron to the k^{th} hidden neuron and f_a is the activation function.

Fig. 3 The flowchart of proposed computational intelligence model



Similarly, for the output of the l^{th} output neuron of CVNN, we have

$$\hat{y}_l = f_a \left(\sum_{k=1}^h V_{lk}^1 z_h^k \right); l = 1, 2, \dots, s \quad (8)$$

In the above equation, V_{lk}^1 is a complex weight that connects the hidden k^{th} neuron to the output l^{th} neuron and f_a is the activation function. Here, the objective of the CVNN is the proper estimation of parameters V^1 and V^0 to minimize the error of the following relationship:

$$E = \frac{1}{2} e^H e = \frac{1}{2} \sum_{k=1}^n e_k \bar{e}_k \quad (9)$$

In the above equation, $e_k = y_k - y_k^A$ and \bar{e}_k is conjugate complex. H is a complex Hermitian operator. The back

propagation algorithm with complex values is used for estimating parameters v^0 and v^1 , as described in [33] in detail. The rule of updating the output weight connecting the hidden k^{th} neuron and the output l^{th} neuron will be as follows:

$$\Delta v_{lk} = \eta_v \delta_l \bar{z}_h^k \quad (10)$$

In this relationship, $\delta_l = y_l - y_l^A$ and η_v is the learning rate that can be a real or complex value. Similarly, the first derivative of the mean square error (MSE) with respect to the input weights $v_{kj}^0; k = 1, \dots, h; j = 1, \dots, m$ is applied to update the input weight connecting the j^{th} input neuron and k^{th} hidden neuron.

$$\Delta v_{kj}^0 = \eta_v \delta_l \left(\sum_{l=1}^n \bar{v}_{lk} \bar{f}'_a \left(\sum_{j=1}^m v_{kj}^0 z_j \right) \right) \bar{z}_j; k = 1, \dots, h, j = 1, \dots, m \quad (11)$$

Inputs-Output Definition & Initialization:

DS : Dataset containing the training / testing samples (Statistical Society = 822 Patients)
CVNN - SIZE : CVNN size of dimension 1×3
CVNN - SIZE (1) = Number of input neurons (m); *CVNN - SIZE (2)* = Number of hidden layer neurons (h); *CVNN - SIZE (3)* = Number of output neurons (s) ($m = 5, h = 10, s = 1$)
 v_{jk}^0 : Input weights of dimension $h \times m$ ($j = 1 \dots m, k = 1 \dots h$)
 v_{lk}^1 : Output weights of dimension $s \times h$ ($l = 1 \dots s, k = 1 \dots h$)
 η : Learning rate parameter ($\eta = 0.001$)
 $K = [k_1, k_2]$: Multiplication constants for logarithm magnitude error and phase error ($k_1 = 1.5, k_2 = 1.5$)
 Υ : Number of epochs ($\Upsilon = 100$)
 Γ : CVNN error threshold ($\Gamma = 0.03$)
 Y : Medical diagnosis of benign or malign mass (real outputs or supervisor samples of *DS*)
 Y^A : Detected benign or malign mass by CVNN (Predicted outputs)

Pre-processing:

For each dataset DS_O , where $O = \{0, 1\}$, representing benign and malign breast mass:
 Do for each $i = 1 \dots n$ // $n =$ all 822 patients
 Normalize each sample (its features values) between $[0, 1]$:
 $X'_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$ // $x_{\min} = 0, x_{\max} = 1$
 End

Stage 1 - Unsupervised Learning Phase:

1.1. For each dataset DS_O , eliminate 'O' (the output or supervisor samples) and create P clusters using SOM (each cluster includes patients with the highest similarity):
 Do for each $i = 1 \dots n$ // $n =$ all 822 patients
 Cluster each sample to $\{C_1, C_2, C_3, C_4\}$:
 1.1.a. $i^*(t) = \arg \min_{v_i} \|x(t) - W_i(t)\|$ // Competition Phase
 1.1.b. $h(i^*, i; t) = \exp \left(-\frac{\|r_i(t) - r_{i^*}(t)\|^2}{2\sigma^2(t)} \right)$ // Cooperation Phase
 1.1.c. $W_i(t+1) = W_i(t) + \eta(t) h(i^*, i; t) [x(t) - W_i(t)]$ // Adaptation Phase
 End

Stage 2 - Supervised Learning Phase (Apply to CVNN training and testing parts, separately):

2.1. For each cluster, add the corresponding output samples which was removed in the previous phase:

Fig. 4 Pseudo-code of the proposed computational intelligence model

```

Do for  $c = 1 \dots P$  //  $P =$  cluster no. obtained from the
previous phase
  Do for each  $i = 1 \dots n$  //  $n =$  all 822 patients
  Transfer sample value from real space  $(R)[0,1]$  to complex space  $(C)[0, \Pi/2]$ :

  2.1.a.  $\phi_i = \frac{\Pi(x_i - a)}{b - a}$  //  $x_i \in [0,1], a = 0, b = 1$ 
  2.1.b.  $f_{R \rightarrow C} = z_j = |r|e^{j\phi} = \cos(\phi) + j \sin(\phi)$  //  $z_i \in [0, \Pi/2], |r| = 1$ 
  End
End

2.2. For each cluster  $P$  (fully complex-valued cluster), Apply CVNN:
Do for  $c = 1 \dots P$ 
  Do for each  $i = 1 \dots n$  //  $n =$  all cluster samples
  While  $(Y > 0)$ 
    2.2.a. Compute hidden layer output:
    
$$z_h^k = \exp\left(\sum_{j=1}^m V_{kj}^0 z_j\right); k = 1, 2, \dots, h$$

    2.2.b. Compute output layer output (predicted value):
    
$$y_l^\wedge = \exp\left(\sum_{k=1}^h V_{lk}^1 z_h^k\right); l = 1, 2, \dots, s$$

    2.2.c. Compute magnitude and phase error for the output (predicted value):
    
$$E = \frac{1}{2} \left( k_1 \log \left[ \frac{\|y\|}{\|y^\wedge\|} \right]^2 + k_2 [\arg(y) - \arg(y^\wedge)]^2 \right)$$

    2.2.d. Compute output delta using the logarithmic error function
    
$$\delta_l = \left( k_1 \log \left[ \frac{|y_l|}{|y_l^\wedge|} \right] + ik_2 \angle y_l - \angle y_l^\wedge \right)$$

    Do for each  $l = 1 \dots s$ 
      Do for each  $k = 1 \dots h$ 
    2.2.e. Update output weight:
    
$$\Delta v_{lk}^1 = \eta z_h^k \delta_l$$

    End
  End

```

Fig. 4 (continued)

In the above equation, \overline{f}'_a is the derivative of the conjugate of the activation function.

2.2.5 Transferring problem space with complex values to that with real values

In the previous section, the phase conversion was carried out in order to transfer the problem space from real values to

complex values considering the complex number magnitude as a constant value. Therefore, the following relationship can be used to convey complex values to the real initial values:

$$f_{C \rightarrow R} = \text{Arc tan} \left(\frac{y}{x} \right) / \frac{\Pi}{2} \tag{12}$$

In the above equation, y and x are the imaginary and real parts of the complex number, respectively. The phase value (ϕ) of the complex number is calculated through calculating

```

End
Do for each  $k = 1 \dots h$ 
    Do for each  $j = 1 \dots m$ 
        2.2.f. Update Input weight:
        
$$\Delta v_{jk}^0 = \eta \left( \sum_{l=1}^{s-1} \overline{v_{lk}} \left( \exp \left( \sum_{j=1}^m v_{kj}^0 z_j \right) \right) \right) \delta_l \bar{z}_j$$

        End
    End
     $\Upsilon = \Upsilon - 1$ 
End
End
End
End
End

3. If ( $E > \Gamma$ )
    Modify changeable values in CVNN ( $\eta, h, k_1, k_2$ ) and go to step 2.2
Else
    Finish.

4. For each predicted sample (detected benign or malign mass), transfer sample value from complex
space ( $C$ )  $[0, \Pi/2]$  to real space ( $R$ )  $[0,1]$ :
    Do for each  $o = 1 \dots n$  //  $n =$  all predicted output samples

    
$$\Theta_o = \text{Arc tan} \left( \frac{y_o}{x_o} \right) // z_o = x_o + jy_o$$

    
$$f_C \rightarrow R = \hat{Y}_o = \Theta_o / \frac{\Pi}{2}$$


End
    
```

Fig. 4 (continued)

$\arctan(\frac{y}{x})$. Furthermore, given that the features of the input space are mapped from $x \in [a, b]$ to $\phi \in [0, \frac{\Pi}{2}]$ by converting real values to the complex values, the real value corresponding with the complex number is obtained by dividing it to $\frac{\Pi}{2}$.

2.3 The proposed computational intelligence model (SOM-CVNN)

In this paper, after the normalization of 822 input samples each with five features and applying unsupervised learning

technique (SOM), separate clusters were defined. It should be mentioned that the determined clusters include patients with a maximum degree of similarity. In the following, the sample values of each cluster are applied to supervised learning technique (CVNN) after transferring to the complex space and separating training and testing samples (693 samples for training and 129 samples to testing the model). In this section, the selection of the unique structure of neural network consisting of neurons number in a hidden layer, choosing the learning rate parameter, and giving initial weights to each of the cluster samples are very important. Then, if the accuracy

Fig. 5 Decision plots for a training phase with 693 patients and b testing phase with 129 patients

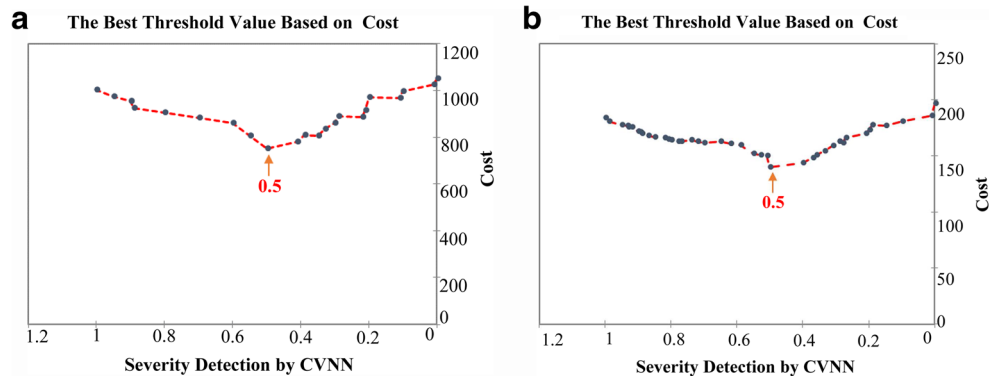
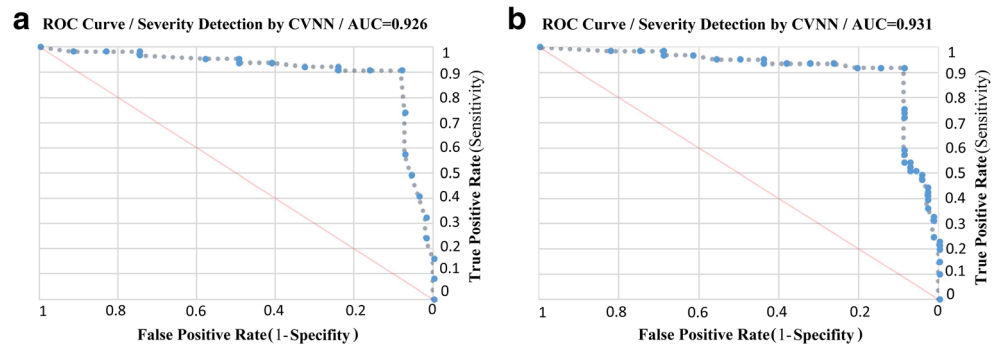


Fig. 6 Graph of the ROC analysis for **a** training data (693 patients) and **b** training data (129 patients)



of the adopted neural network during the training process is satisfying (magnitude and phase error obtained for the complex number are a minimum value), the testing samples would be applied to the network. Otherwise, the previous step is repeated, the design of neural network architecture is modified, and efforts are made to reduce the error. In the end, the complex output obtained from the network is transferred to real space and the evaluation criteria of the confusion matrix and receiver operating characteristic (ROC) are considered for the total samples of clusters. The overall process of the proposed algorithm is shown in the flowchart of Fig. 3.

2.3.1 The proposed structure of SOM for patient clustering

In choosing the optimal size of the SOM (number of neurons = number of clusters), there is no theoretical and appropriate base, and the size is determined according to the considered application [34]. In this paper, considering the small number of input samples (822 samples), the size of the SOM is considered to be a 2 × 2 neuron network, and consequently, the number of clusters will be 4.

2.3.2 The proposed structure of CVNN for breast cancer detection

Here, the intention is to use the exponential function as an activation function for processing non-linear complex data and a logarithmic error function including magnitude and phase error with complex values. Moreover, back propagation

(BP) learning algorithm with complex values was used for CVNN training. BP learning algorithm with complex values based on the logarithmic error function is quite similar to BP learning algorithm based on the mean square error.

In the following, the pseudo-code of the proposed computational model is coded step by step (Fig. 4).

2.4 Statistics (confusion matrix and ROC curve analysis)

In statistics, a ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. TPR is also known as sensitivity, recall, or probability of detection in machine learning [35]. FPR is also known as the fallout or probability of false alarm and can be calculated as (1 – specificity). The ROC curve is, thus, sensitivity as a function of fallout. In general, if TPR and FPR are known, the ROC curve can be generated by plotting the cumulative distribution (area under the probability distribution from $-\infty$ to the discrimination threshold) function of TPR on the y axis versus the cumulative distribution function of the FPR on x axis.

In this article, a diagnostic test that seeks to determine whether a patient has a cancerous or non-cancerous mass (benign or malignant mass) was taken into consideration. There are four possible outcomes from a binary classifier. If the outcome from a detective model is **P** and the actual value from a medical diagnosis test is also **P**, then it is called a true positive

Table 1 Confusion matrix for training samples (693 patients in four clusters)

		The best threshold value = 0.5		
		Detected (predicted) condition		
		Malign	Benign	Total
Actual condition	Malign	TP = 314	FN = 23	337
	Benign	FP = 20	TN = 336	356
	Total	334	359	693

Table 2 Confusion matrix for testing samples (129 patients in four clusters)

		The best threshold value = 0.5		
		Detected (predicted) condition		
		Malign	Benign	Total
Actual condition	Malign	TP = 58	FN = 3	61
	Benign	FP = 4	TN = 64	68
	Total	62	67	129

Table 3 Comparison between proposed models (obtained results in testing phase)

Author (s)	Year	Method	Sensitivity	Specificity	Accuracy	ROC AUC	Patient no. ^a
Saritas [4]	2012	ANN	0.902	0.814	0.855	0.87	822
Nugroho Et al. [15]	2013	NB-SMO	0.81	0.84	0.833	0.9	822
Our study	2017	SOM-CVNN	0.95	0.94	0.945	0.93	822

^a All of the methods have used the dataset prepared by [22] with 822 patients

(TP); however, if the actual value is N , then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the detective model outcome and the actual value are N , and false negative (FN) is when the detective model outcome is N , while the actual value is P . A FP in this case occurs when the person tests positive, but does not actually have the disease. A FN, on the other hand, occurs when the person tests negative, suggesting that they are healthy, when they actually do have the malign mass. An experiment was defined from P positive patients and N negative patients for some condition. The above definitions are as follows:

$$\begin{aligned} \text{TPR} &= \text{Sensitivity} \\ &= \Sigma \text{ True positive (TP)} / \Sigma \text{ Condition positive (TP + FN)} \end{aligned} \quad (13)$$

$$\begin{aligned} \text{TNR} &= \text{Specificity} \\ &= \Sigma \text{ True negative (TN)} / \Sigma \text{ Condition negative (TN + FP)} \end{aligned} \quad (14)$$

$$\text{FPR} = 1 - \text{TNR (sensitivity)} \quad (15)$$

ROC analysis is often called the ROC accuracy ratio, a common technique for judging the accuracy of default probability models. ROC curves are widely used in laboratory medicine to assess the diagnostic accuracy of a test, to choose the optimal cutoff of a test, and to compare diagnostic accuracy of several tests. Thus, we use accuracy criterion with the following formula:

$$\begin{aligned} \text{Accuracy} &= \Sigma \text{ True positive (TP)} \\ &+ \Sigma \text{ True negative (TN)} / \Sigma \text{ Total population (P + N)} \end{aligned} \quad (16)$$

An ROC space is defined by FPR and TPR as x and y axes, respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). Since TPR is equivalent to sensitivity and FPR is equal to $1 - \text{specificity}$, the ROC graph is sometimes called the sensitivity versus ($1 - \text{specificity}$) plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space. The best possible prediction method would yield a point in the upper left corner or coordinate (0, 1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). The (0, 1) point is also called a perfect

classification. In addition, the diagonal divides the ROC space. The points above the diagonal represent good classification results, and the points below the line represent poor results.

3 Results

In this article, the detected outputs of the CVNN change between [0, 1]. These outputs are defined as thresholds. ROC analysis displays for each possible threshold value, the value of the various performance indices. We wish to strongly penalize diagnostic errors and more particularly the case where the sick patients are not detected (FN). So, for each threshold, different cost values are assigned to the TP, TN, FP, and FN outcomes. For example, if we decide to detect a patient's mass status when the mass severity is greater than or equal to 0.9, the sensitivity is 0.31, the specificity is 0.98, and the total cost is 172. In fact, we made a decision to select an appropriate threshold based on total cost values. For this purpose, the values assigned to TP, TN, FP, and FN were 1, 1, 2, and 4, respectively. As can be seen in Fig. 5, the decision plot allows to choose the threshold value that minimizes the cost. This value corresponds to a severity of 0.5, for both training and testing sections.

Thus, according to Fig. 6, based on ROC analysis, the area below CVNN training and testing data was obtained to be 0.92 and 0.93, respectively.

4 Discussion

For training and testing phases, the four outcomes obtained from CVNN are shown in Tables 1 and 2 as a 2×2 confusion matrix.

According to the test dataset confusion matrix analysis, it is observed that the ratio of distinguishing the sick from the true sick is 95%, the ability to distinguish the healthy from the true healthy is 94.2%, and the ratio of the total number of correct diagnoses of the test as sick and healthy is 94.5%. It is obvious that the proposed model that was applied was successful in breast cancer detection, in a reliable manner. Additionally, for comparison of several proposed models in this field, the

sensitivity, specificity, and AUC are also apt criteria to compare different diagnostic tests. Thus, according to Table 3, it can be observed that the current work is completely successful in comparison with other proposed diagnostic methods.

5 Conclusion

This paper proposed a hybrid computational intelligence model, namely, SOM-CVNN for diagnosis of breast cancer. We applied a real dataset including 822 patients with five features. The SOM technique was used to cluster the patients, and then for each cluster, the patient's features were applied to complex-valued neural network and dealt with to classify breast cancer severity (benign or malign). In the testing phase, health and disease detection ratios were 94 and 95%, respectively. The obtained results show that the model can be used as a reliable tool that may eliminate unnecessary biopsy. Furthermore, better breast cancer detection results can be obtained by increasing the number of patients in conjunction with other innovative machine learning methods. We are working on another novel hybrid model which is extremely dependent on data mining along with pattern recognition.

Acknowledgements The authors would like to specially thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the article.

Funding There was no funding associated with this research.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

Research involving human participants and/or animals This article does not contain any studies with human participants performed by any of the authors.

References

- World Cancer Report. (2014). <http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014>
- Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F (2013) Cancer Incidence and Mortality Worldwide: IARC CancerBase No 11. <http://globocan.iarc.fr>
- Saritas I, Ozkan IA, Sert IU (2010) Prognosis of prostate cancer by artificial neural networks. *Expert Syst Appl* 37(9):6646–6650. <https://doi.org/10.1016/j.eswa.2010.03.056>
- Saritas I (2012) Prediction of breast cancer using artificial neural networks. *J Med Syst* 36(5):2901–2907. <https://doi.org/10.1007/s10916-011-9768-0>
- Saad G, Khadour A, Kanafani Q (2016) ANN and Adaboost application for automatic detection of microcalcifications in breast cancer. *Egypt J Radiol Nucl Med* 47(4):1803–1814. <https://doi.org/10.1016/j.ejmm.2016.08.020>
- Groshev A (2016) Chapter 18—recent advances of biochemical analysis: ANN as a tool for earlier cancer detection and treatment A2—Puri, Munish. In: Pathak Y, Sutariya VK, Tipparaju S, Moreno W (eds) *Artificial neural network for drug design, delivery and disposition*. Academic Press, Boston, pp 357–375. <https://doi.org/10.1016/B978-0-12-801559-9.00018-1>
- Wahab N, Khan A, Lee YS (2017) Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. *Comput Biol Med* 85:86–97. <https://doi.org/10.1016/j.compbimed.2017.04.012>
- Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Appl* 36(2, Part 2):3465–3469. <https://doi.org/10.1016/j.eswa.2008.02.064>
- Huang M-L, Hung Y-H, Chen W-Y (2010) Neural network classifier with entropy based feature selection on breast cancer diagnosis. *J Med Syst* 34(5):865–873. <https://doi.org/10.1007/s10916-009-9301-x>
- Senapati MR, Mohanty AK, Dash S, Dash PK (2013) Local linear wavelet neural network for breast cancer recognition. *Neural Comput & Applic* 22(1):125–131. <https://doi.org/10.1007/s00521-011-0670-y>
- Ludwig SA, Roos S (2010) Prognosis of breast cancer using genetic programming. In: Setchi R, Jordanov I, Howlett RJ, Jain LC (eds) *Knowledge-based and intelligent information and engineering systems: 14th international conference, KES 2010, Cardiff, UK, September 8–10, 2010, proceedings, Part IV*. Springer, Berlin Heidelberg, pp 536–545. https://doi.org/10.1007/978-3-642-15384-6_57
- Chao C-M, Y-W Y, Cheng B-W, Kuo Y-L (2014) Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree. *J Med Syst* 38(10): 106. <https://doi.org/10.1007/s10916-014-0106-1>
- Shin HJ, Kim HH, Shin KC, Sung YS, Cha JH, Lee JW, Son BH, Ahn SH (2016) Prediction of low-risk breast cancer using perfusion parameters and apparent diffusion coefficient. *Magn Reson Imaging* 34(2):67–74. <https://doi.org/10.1016/j.mri.2015.10.028>
- Korkmaz SA, Korkmaz MF, Poyraz M (2016) Diagnosis of breast cancer in light microscopic and mammographic images textures using relative entropy via kernel estimation. *Med Biol Eng Comput* 54(4):561–573. <https://doi.org/10.1007/s11517-015-1361-0>
- Nugroho KA, Setiawan NA, Adji TB (2013) Cascade generalization for breast cancer detection. In: *Information Technology and Electrical Engineering (ed) (ICITEE), 2013 International Conference on*. IEEE, New York, pp 57–61
- Meng L, Xu Y, Xu C, Zhang W (2016) Biomarker discovery to improve prediction of breast cancer survival: using gene expression profiling, meta-analysis, and tissue validation. *Oncotargets Ther* 9: 6177
- Ding X, Cao J, Alsaedi A, Alsaadi FE, Hayat T (2017) Robust fixed-time synchronization for uncertain complex-valued neural networks with discontinuous activation functions. *Neural Netw* 90:42–55. <https://doi.org/10.1016/j.neunet.2017.03.006>
- Song Q, Shu H, Zhao Z, Liu Y, Alsaadi FE (2017) Lagrange stability analysis for complex-valued neural networks with leakage delay and mixed time-varying delays. *Neurocomputing* 244:33–41. <https://doi.org/10.1016/j.neucom.2017.03.015>
- Wang L, Song Q, Liu Y, Zhao Z, Alsaadi FE (2017) Global asymptotic stability of impulsive fractional-order complex-valued neural networks with time delay. *Neurocomputing* 243:49–59. <https://doi.org/10.1016/j.neucom.2017.02.086>
- Borkowska EM, Kruk A, Jedrzejczyk A, Rozniecki M, Jablonowski Z, Traczyk M, Constantinou M, Banaszkiwicz M,

- Pietrusinski M, Sosnowski M, Hamdy FC, Peter S, Catto JWF, Kaluzewski B (2014) Molecular subtyping of bladder cancer using Kohonen self-organizing maps. *European Urology Supplements* 12(4):e1213–C1105. [https://doi.org/10.1016/S1569-9056\(13\)61953-3](https://doi.org/10.1016/S1569-9056(13)61953-3)
21. Borkowska EM, Kruk A, Jedrzejczyk A, Rozniecki M, Jablonowski Z, Traczyk M, Constantinou M, Banaszkiwicz M, Pietrusinski M, Sosnowski M, Hamdy FC, Peter S, Catto JW, Kaluzewski B (2014) 995 Kohonen's self-organizing map for molecular subtyping in bladder cancer. *European urology supplements* 13(1):e995. [https://doi.org/10.1016/S1569-9056\(14\)60978-7](https://doi.org/10.1016/S1569-9056(14)60978-7)
 22. Elter M, Schulz-Wendtland R, Wittenberg T (2007) The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. *Med Phys* 34(11): 4164–4172
 23. Santen RJ (2014) *Benign breast disease in women*. MDText.com, Inc, South Dartmouth (MA)
 24. McKenna RJ (1994) The abnormal mammogram radiographic findings, diagnostic options, pathology, and stage of cancer diagnosis. *Cancer* 74(S1):244–255. <https://doi.org/10.1002/cncr.2820741308>
 25. Boyd NF, Martin LJ, Rommens JM, Paterson AD, Minkin S, Yaffe MJ, Stone J, Hopper JL (2009) Mammographic density: a heritable risk factor for breast cancer. In: Verma M (ed) *Cancer epidemiology: modifiable factors*. Humana Press, Totowa, NJ, pp 343–360. https://doi.org/10.1007/978-1-60327-492-0_15
 26. Gülsün M, Demirkazık FB, Köksal A, Arıyürek M (2002) According to BI-RADS assessment of breast microcalcifications and to investigate the agreement between reviewers. *Off J Turkish Soc Radiol* 8(3):358–363
 27. Edward T, Rick K, Robert R (2011) *Conn's current therapy*. Elsevier INC. Saunders, Philadelphia
 28. Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* 43(1):59–69
 29. Douzas G, Bacao F (2017) Self-organizing map oversampling (SOMO) for imbalanced data set learning. *Expert Syst Appl* 82: 40–52. <https://doi.org/10.1016/j.eswa.2017.03.073>
 30. Novaes CG, Romão ILS, Santos BG, Ribeiro JP, Bezerra MA, da Silva EGP (2017) Screening of *Passiflora L.* mineral content using principal component analysis and Kohonen self-organizing maps. *Food Chem* 233:507–513. <https://doi.org/10.1016/j.foodchem.2017.04.111>
 31. Shirazi AZ, Mohammadi Z (2016) A hybrid intelligent model combining ANN and imperialist competitive algorithm for prediction of corrosion rate in 3C steel under seawater environment. *Neural Comput & Applic*:1–10
 32. Shirazi AZ, Hatami M, Yaghoobi M, Chabok SJS (2016) An intelligent approach to predict vibration rate in a real gas turbine. *Intell Ind Syst* 2(3):253–267
 33. Suresh S, Sundararajan N, Savitha R (2013) Erratum: supervised learning with complex-valued neural networks. In: *Supervised learning with complex-valued neural networks*. Springer, Berlin, pp E1–E1
 34. Chang L-C, Shen H-Y, Chang F-J (2014) Regional flood inundation nowcast using hybrid SOM and dynamic neural networks. *J Hydrol* 519:476–489
 35. Saito T, Rehmsmeier M (2016) Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics*:btw570



His research interests include bioinformatics and application of machine learning and AI techniques in medicine.

Amin Zadeh Shirazi He received the bachelor degree in Hardware Engineering and the master degree in Artificial Intelligence from Khayyam University of Mashhad and Islamic Azad University of Mashhad, Iran, in 2010 and 2014, respectively. He is currently a lecturer in computer department at University of Applied Science and Technology (UAST). He has more than 15 published articles in well-known journals and conferences.



soft computing in industry and medicine. To date, he has published more than 40 articles in numerous conferences and journals.

Seyyed Javad Seyyed Mahdavi Chabok He obtained his Bachelor of Science with first class honors from the Sharif University, Iran. He then pursued his Ph.D. at the Iran University of Science and Technology, where he graduated with distinction and commendation. In 2006, he joined the Islamic Azad University of Mashhad as a senior lecturer, and was subsequently promoted to associate professor in 2016. His current research areas include the application of



Zahra Mohammadi She received her B.Sc., Computer Hardware Engineering degree from Khayyam University of Mashhad, Iran, in 2010. Also, she obtained M.Sc., in Computer Engineering (artificial intelligence) degree from Imam Reza International University, Mashhad, Iran, 2015. Currently, she works as a lecturer at the same university. Her main research interest includes neural networks and machine learning.