Research Article

Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier

C. V. Subbulakshmi and S. N. Deepa

Department of EEE, Anna University Regional Centre, Coimbatore, Coimbatore 641 047, India

Correspondence should be addressed to C. V. Subbulakshmi; cvs.lakshmi@gmail.com

Medical data classification is a prime data mining problem being discussed about for a decade that has attracted several researchers around the world. Most classifiers are designed so as to learn from the data itself using a training process, because complete expert knowledge to determine classifier parameters is impracticable. This paper proposes a hybrid methodology based on machine learning paradigm. This paradigm integrates the successful exploration mechanism called self-regulated learning capability of the particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) classifier. As a recent off-line learning method, ELM is a single-hidden layer feedforward neural network (FFNN), proved to be an excellent classifier with large number of hidden layer neurons. In this research, PSO is used to determine the optimum set of parameters for the ELM, thus reducing the number of hidden layer neurons, and it further improves the network generalization performance. The proposed method is experimented on five benchmarked datasets of the UCI Machine Learning Repository for handling medical dataset classification. Simulation results show that the proposed approach is able to achieve good generalization performance, compared to the results of other classifiers.

1. Introduction

In recent times, the application of computational or machine intelligence in medical diagnosis is a new trend for large medical data applications. Most of the diagnosis techniques in medical field are systematized as intelligent data classification approaches. In computer-aided decision (CAD) systems, information technology methods are adopted to assist a physician to diagnose disease of a patient. Among the various assignments performed by a CAD system, classification is most common, where a tag is allocated to a query case (i.e., a patient) based on chosen number of features (i.e., medical findings). Thus medical database classification problem may be categorized as a class of complex optimization problem with an objective to guarantee the diagnosis aid accurately. Aside from other traditional classification problems, medical dataset classification problems are also applied in future diagnosis. Generally, patients or doctors are not completely informed about the cause (classification result) of the disease, but also will be made known of the symptoms that derive the cause of disease, which is the most important of the medical dataset classification problem.

In a classification problem, the objective is to learn the decision surface that accurately maps an input feature space to an output space of class labels [1]. In medical field, various computer researchers have attempted to apply diverse techniques to improve the accuracy of data classification for the given data, classification techniques whose classification accuracy will better yield enough information to identify the potential patients and thereby improvise the diagnosis accuracy. In the recent studies, metaheuristic algorithms (to name a few, simulated annealing, genetic algorithms, and particle swarm optimizations) and also data mining techniques (to name a few, Bayesian networks, artificial neural network, fuzzy logic, and decision tree) were applied for classification of medical data and obtained with remarkably meaningful results.

As per literature and above discussion, for past decades several classification tools are available for medical dataset classification. Even then, artificial neural networks (ANNs) are widely accepted and utilized to solve the real world classification problems in clinical applications. Because of their generalization and conditioning capabilities, requirement of minimal training points, and faster convergence time, ANNs are found to perform better and result in faster output in comparison with the conventional classifiers [2]. Various learning/training algorithms for several neural network architectures have been proposed in various problems in science, engineering, and technology and even in some parts of business industry and medicine. A few notable classification applications include the handwritten character recognition, speech recognition, biomedical medical diagnosis, text categorization, information retrieval, and prediction of bankruptcy [3, 4].

Several versions of ANN are modeled, to name a few, feedforward neural network, Boltzmann machine, radial basis function (RBF) network, Kohonen self-organizing network, learning vector quantization, recurrent neural network, Hopfield network, spiking neural networks, and extreme learning machine (ELM); most of them are inspired by biological neural networks. Among these several ANNs, feedforward neural networks (FFNNs) are popularly and widely used for the classification/identification of linear/nonlinear systems. To prevail over the slow construction of FFNN models, several new training schemes were introduced and amongst them the extreme learning machine (ELM) has gained wide attention in recent days [4]. One of the very unusual individualities of this ELM is its nontraditional training procedure. Here, ELM randomly selects the input layer and hidden layer connection weights and also estimates the connection weights between the hidden layer and output layer analytically. Besides, the ELM tends to require more hidden neurons compared to conventional tuning-based learning algorithms; perhaps this has been believed to be trivial compared to other positive features of ELM [5].

In spite of its superiority with other FFNN training algorithms, various advancements have been proposed to improvise its performance in the last few years, merely by hybridizing the ELM with recent metaheuristic algorithms. Hybridization was done with two ideas: one is a feature selection approach using the ELM as wrapper classifier [6] and the other theme is using microevolution in order to obtain the best set of weights and biases in the input layer of the ELM [7]. This research opts for the second category, where one of the very effective metaheuristic algorithms has been used jointly with analytical methods for FFNN training. With a growing body of literature in this category of hybridization for training the ELM and applications of classifiers for medical set classification, a comprehensive list of a few such literatures is given in [7–26].

In [27] a new nonlinear system identification scheme is proposed, where differential evolution (DE) is used to optimize the initial weights used by a Levenberg-Marquardt (LM) algorithm in the learning of a FFNN. In [28] a similar method was proposed using a simulated annealing (SA) approach and subsequently the training of FFNN with a back propagation method, which is computationally expensive. In [29], the process of selecting the input weights and hidden biases, using an improved PSO and validation set of the output weights as well as constrained input weights and hidden biases, was within a reasonable range. Again in [30] the coral reefs optimization (CRO) has been used for fast convergence to optimal values and had been used for carrying out evolution in ELM weights, in order to enhance the performance of these machines.

In this paper, again a PSO based hybrid algorithm to train the ELM is proposed. The proposed hybrid algorithm is modeled in such a way that the ELM solves the main problem, whereas the PSO evolves the weights of the FFFN, to improve the solutions obtained and further to improve the network generalization performance. In order to obtain a robust SLFN, otherwise the ELM for classification accuracy, the PSO algorithm is enhanced by incorporating a velocity updating mechanism which diversifies the search process of the particles and could escape all local traps and guarantees much better solutions for the ELM. The proposed hybrid model is able to classify some of the UCI medical datasets, namely, Wisconsin Breast Cancer, Pima Indians Diabetes, Heart-Statlog, Hepatitis, and Cleveland Heart Disease. The medical data classification was carried out by extracting and analyzing available data with a suitable sampling procedure.

This paper is organized as follows. The FFFN architecture for ELM is overviewed in Section 2. Section 3 gives a brief review of the ELM. The proposed improved particle swarm optimization (PSO) algorithm is presented in Section 4. Section 5 gives a detailed formulation of the proposed optimization methods for the ELM. Section 6 presents experimental results. Finally, concluding remarks are drawn in Section 7.

2. ELM for Medical Dataset Classification

This research for medical data classification relies on the performance of the extreme learning machine (ELM) classifier proposed in [4], which handles the training for single-hidden layer feedforward neural networks. An introduction to the ELM will be presented in the next section.

2.1. Extreme Learning Machine Classifier. The extreme learning machine (ELM) was originally developed in 1992 [3, 4] and can be categorized as a supervised learning algorithm capable of solving linear and nonlinear classification problems. When compared to other neural networks architectures, ELM may be understood as a single layer feedforward neural net (FFNN) with one hidden layer. The prime constituting blocks of ELMs are structural risk minimization, originating from statistical learning theory, nonlinear optimization, and duality and kernel induced features spaces, underlining the technique with an exact mathematical framework.

ELM [5, 6] is best suited for larger training samples and also the effect of number of hidden neurons using different ratios of the number of features of testing and training data was examined. This classifier is compared with that of the
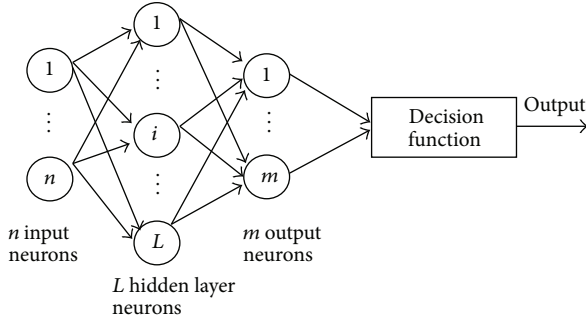
FIGURE 1: Architecture of the ELM (single layer FFNN).

conventional neural network classifiers using the classification rate for medical data classification. The main motivation of the extreme learning machine classification is to separate classification data with a linear decision surface and maximize the margin in between the various categories of classes. This leads to the convex quadratic programming problem. It can be seen that training the ELM involves solving a quadratic optimization problem which requires the use of optimization routines from various recent mathematical or heuristic approaches. Specifically ELM classifier is chosen for considered application due to the following facts: extreme learning machine classifier provides good solutions, even for difficult search spaces and when complex datasets are used as in this case, extreme learning machine (ELM) is found to be a competitive good solution provider due to its converging characteristics. Further, extreme learning machine classifier reduces the computational burden and time prevailed in earlier classifiers. ELM achieves good generalization performance at extremely fast learning speed. Figure 1 shows the basic ELM architecture.

The basic ELM classifier algorithm is given as follows.

Given N different samples $(x_i, y_i) \in R^n \times R^m$, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ is the input vector and $y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^T$ is the target vector, standard SLFNs with M hidden nodes and activation function $g(x)$ are formulated as

$$\sum_{i=1}^{M} \alpha_i f_i(x_j) = \sum_{i=1}^{M} \alpha_i f(w_i \cdot x_j + b_i) = O_j, \quad j = 1, \ldots, N,$$

(1)

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the input nodes and the ith hidden nodes, $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{im}]^T$ is the weight vector connecting the ith hidden node and the output nodes, and b_i is the bias of the ith hidden node. Assume that the function approaches all N samples by zero error; that is, there exist parameters (w_i, b_i) and α_i such that

$$\sum_{i=1}^{M} \alpha_i f(w_i \cdot x_j + b_i) = y_j, \quad j = 1, \ldots, N.$$

(2)

The N equations can be simplified as $H\alpha = T$, where

$$H(w_1, \ldots, w_M, b_1, \ldots, b_M, x_1, \ldots, x_N)$$
$$= \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \cdots & f(w_M \cdot x_1 + b_M) \\ \vdots & \ddots & \vdots \\ f(w_1 \cdot x_N + b_1) & \cdots & f(w_M \cdot x_N + b_M) \end{bmatrix}_{N \times M},$$
$$\alpha = [\alpha_1^T, \ldots, \alpha_M^T]_{M \times m}^T, \quad T = [y_1^T, \ldots, y_N^T]_{N \times m}^T.$$

(3)

The solution of the linear system is $\alpha = H^\dagger T$, where $H^\dagger$ is the Moore-Penrose generalized inverse of hidden layer output matrix H. ELM algorithm can be written as the following three steps.

> Input: it is a training dataset $\{(x_j, y_j) \mid x_j \in \vert R^n, \ t_j \in R^m, \ j = 1, \ldots, N\}$, an activation function g, and the number of hidden nodes M.
>
> Output: it is the weights of hidden layer to output layer.

(1) Randomly generate the parameters of weights and bias $(w_i, b_i), i = 1, \ldots, M$.

(2) Calculate the hidden layer output matrix H.

(3) Calculate the weights matrix of output $\alpha = H^\dagger T$.

Compared to general artificial neural networks the ELM method proffers a considerably smaller number of parameters for tuning. The main modeling idea consists in the choice of a kernel function and the equivalent kernel parameters have control over the convergence speed and the quality of final solution obtained.

3. Particle Swarm Optimization: An Overview

In information technology era swarm intelligence is the domain that derives its models from natural and artificial systems comprised of many individuals that coordinate using self-organization and decentralized control. One of the swarm intelligence methods, the particle swarm optimization (PSO) algorithm, was first proposed by Kennedy and Eberhart [32, 33]. It is inspired by observations from social dynamics of bird flocking or fish schooling. The idea arises from the natural behavior that a large number of birds flock parallelly, change direction spontaneously, scatter and regroup at intervals, and finally reach a target. This form of social behaviour increases the success rate for food foraging and expedites the process of reaching a target. This PSO algorithm simulating bird foraging or bee hiving activity can be modeled as an optimizer for nonlinear functions of coninuous and discrete variables. Several literatures have already detailed the PSO; hence we just give a simple flowchart of the PSO as in Figure 2.

3.1. PSO Algorithm: Explained. The PSO algorithm searches the solution space using a swarm of n particles (i.e.,
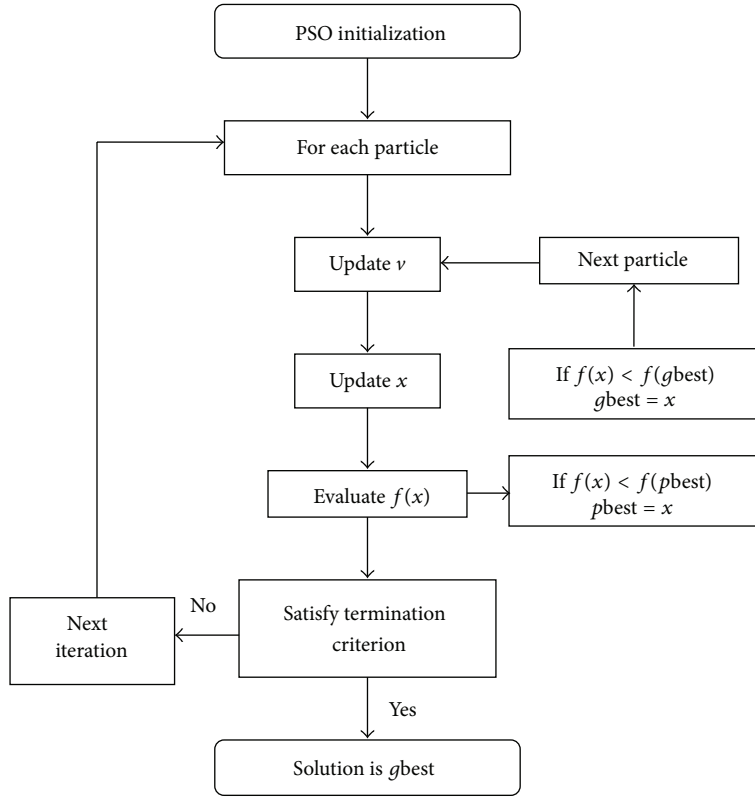
Figure 2: Flowchart of the basic PSO.

$[x^1, x^2, x^3, \ldots, x^n]$), randomly positioned in the search space of the optimization problem. Each particle, which denotes a possible solution, is represented as a position $x_k^i = \begin{bmatrix} x_k^{i,1} & x_k^{i,2} & \cdots & x_k^{i,m} \end{bmatrix}^T$ and a velocity given by $v_k^i = \begin{bmatrix} v_k^{i,1} & v_k^{i,2} & \cdots & v_k^{i,m} \end{bmatrix}^T$. For every algorithmic iteration, the ith particle position evolves using the following update rules:

$$x_{k+1}^i = x_k^i + v_{k+1}^i, \tag{4}$$

$$v_{k+1}^i = \left(w_{k+1}\right) v_k^i + c_1 r_{1,k}^i \left(p_k^i - x_k^i\right) + c_2 r_{2,k}^i \left(p_k^g - x_k^i\right), \tag{5}$$

where w_{k+1} is the linear inertia constant, c_1, c_2 are the acceleration factors for cognitive and social component, respectively, $r_{1,k}^i, r_{2,k}^i$ are random numbers generated in [0, 1], p_k^i is the personnel best position of the ith particle, and p_k^g is the best among all the personnel bests in the entire particles in the current iteration k.

In order to improve the exploration and exploitation capacities of the proposed PSO algorithm, we choose for the inertia factor a linear evolution with respect to the algorithm iteration as given by Shi and Eberhart in [34]:

$$w_{k+1} = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{k_{\max}}\right) k, \tag{6}$$

where $w_{\max} = 0.9$ and $w_{\min} = 0.4$ represent the maximum and minimum inertia factor values, respectively, and $k_{\max}$ is the maximum iteration number. Like the other metaheuristic

method, the PSO algorithm is initially developed as an unconstrained optimizer. Finally, the basic PSO algorithm can be algorithmically understood from the following steps.

(1) The parameters of the PSO to initialize search are to be defined, size n_p, maximum and minimum inertia weight values, c_1 and c_2 mostly equal to 2, and so forth.

(2) Initialize the n_p particles by randomly generating the positions x_0^i from the solution space and velocities v_0^i are generally zero. Evaluate the initial population using the fitness function and determine the personnel best and local best.

(3) Increment the iteration number k. For each particle apply the update equations (4) and (5), and evaluate the corresponding fitness values $\varphi_k^i = \varphi(x_k^i)$:

   (i) if $\varphi_k^i \leq pbest_k^i$ then $pbest_k^i = \varphi_k^i$ and $p_k^i = x_k^i$;

   (ii) if $\varphi_k^i \leq gbest_k$ then $gbest_k = \varphi_k^i$ and $p_k^g = x_k^i$,

   where $pbest_k^i$ and $gbest_k$ represent the best previous fitness of the ith particle and the entire swarm, respectively.

(4) If the termination criterion is satisfied, the algorithm terminates with the final solution. Otherwise, go to step 3.

However, these basic variants of PSO suffer from some common problems, which are quite apparent in several such

stochastic optimization algorithms, for example, "curse of dimensionality" and tendency of premature convergence and hence getting stuck in local optima. Hence, some improved versions of PSOs have been very recently proposed to address some of these specific drawbacks. In this paper, we propose one such improved version of PSO algorithms, called the self-regulated learning PSO algorithm (henceforth called SRLPSO), which attempts to overcome the problem of both "curse of dimensionality" and tendency of premature convergence.

3.2. Mechanism for the Self-Regulated Learning PSO Algorithm. Let us revisit the velocity updating equation as given in (5). Generally the first term in the above equation is the previous controlled velocity. Second term is the cognitive component which is responsible for the particles own knowledge in the search space. The third term is called social component which is the real mechanism that decides the overall exploration and exploitation capacities of the proposed PSO algorithm.

Here p_k^g is the global best in the current iteration. This p_k^g is the position of the particle that will assume values that contribute the best fitness value in the current iteration. We replace p_k^g by P_k^M, where P_k^M will assume values from both best (global) position p_k^g and worst position p_k^W. Thus each component in P_k^M will take a place based on a self-regulated acceptance criterion. Thereby in the beginning of the runs, the majority of the components of P_k^M will be influenced by p_k^W, to establish exploration, and in long run with p_k^g as the algorithm proceeds to its final iterations to establish exploitation. The self-regulating PSO achieves faster convergence in comparison with conventional PSO due to the fact that the global best point will take values from both best position and worst position and each component updation will take a place based on a self-regulated acceptance criterion, thus establishing exploration in the initial iterations and proceeding towards final iterations to establish exploitation. Thereby the equation is rewritten as

$$v_{k+1}^i = (w_{k+1}) v_k^i + c_1 r_{1,k}^i \left(p_k^i - x_k^i\right) + c_2 r_{2,k}^i \left(P_k^M - x_k^i\right),$$
$$P_k^M = \begin{bmatrix} P_k^{i,1} & P_k^{i,2} & \cdots & P_k^{i,m} \end{bmatrix}. \tag{7}$$

The choice of each component in P_k^M is the main contribution in this research work and it is explained as follows:

$$P_k^{i,1} = \begin{cases} p_k^g, & \text{if } \exp(\text{iter}) < 0.5 \\ p_k^W, & \text{otherwise,} \end{cases} \tag{8}$$

where p_k^W is the position of the particle that will assume values that contribute the worst fitness value in the current iteration. iter is the iteration count in the current iteration. This mechanism is proved to be a best diversifier as it is proved as a global mechanism for thoroughly exploring the solution space.

3.3. The Flowchart of the Proposed Self-Regulated Learning PSO (SRLPSO) Algorithm. The flowchart of the self-regulated learning scheme is shown in Figure 3.

4. Proposed SRLPSO Based ELM for Medical Dataset Classification

This proposed methodology combines the concept of SRLPSO for optimizing the weights in ELM neural network. This self-regulated learning PSO with ELM enables the selection of input weights to increase the generalization performance and the conditioning of the single layer feedforward neural network. The steps of the proposed approach are as follows.

Step 1. Initialize positions with a set of input weights and hidden biases: $[W_{11}, W_{12}, \ldots, W_{1n}, \ldots, W_{21}, W_{22}, \ldots, W_{2n}, \ldots, W_{H1}, W_{H2}, \ldots, W_{Hn}, b_1, b_2, \ldots, b_H]$. These will be randomly initialized within the range of $[-1, 1]$ on D dimensions in the search space.

Step 2. For each member in the group, the respective output final weights are computed at ELM as given in (3).

Step 3. Now invoke self-regulated learning PSO as in Section 3.3.

Step 4. Then the fitness, which is the mean square error (MSE) of each member, is evaluated as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} E_i^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_k^i - d_k^i\right)^2, \tag{9}$$

where N is the number of training samples and the terms y_k and d_k are the error of the actual output and target output of the kth output neuron of ith sample. Thus, fitness function $f(x)$ is defined by the MSE. In order to avoid overfitting of the single layer feedforward neural network, the fitness of each member is adopted as the mean squared error (MSE) on the validation set only instead of the whole training set as in [35].

Step 5. Find the SRL acceptance based on the fitness.

Step 6. Update the velocity and position equations of each particle as given in (4) and (5).

Step 7. Stopping criteria: the algorithm repeats Steps 2 to 6 until certain criteria are met, along with hard threshold value as maximum number of iterations. Once stopped, the algorithm reports values with optimal weights with minimal MSE as its solution.

Thus SRLPSO with ELM finds the best optimal weights W and bias b so that the fitness reaches the minimum to achieve better generalization performance, with minimum number of hidden neurons, considering both the advantages of both ELM and SRLPSO. In the process of selecting the input weights, the self-regulated learning PSO considers not only the MSE on validation set but also the norm of the output weights. The proposed SRLPSO based ELM will combine the
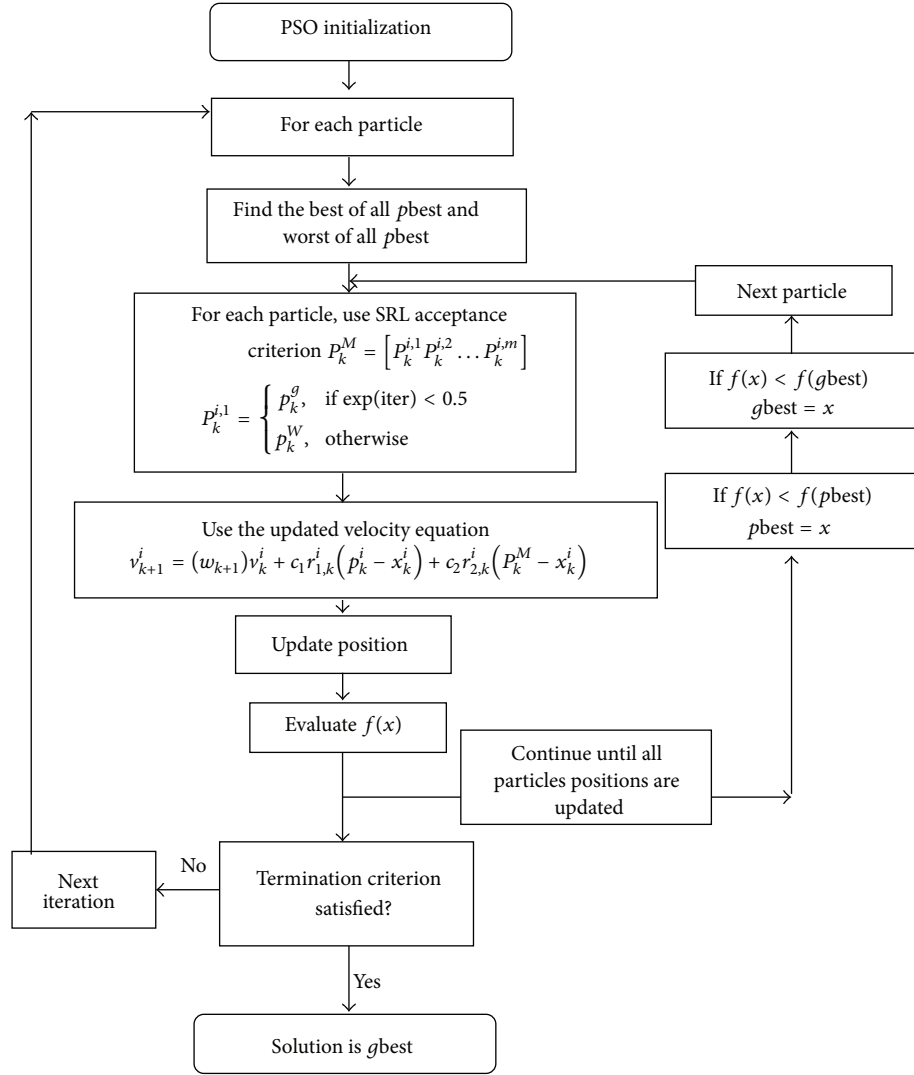
FIGURE 3: Flowchart of the PSO with self-regulated learning scheme.

feature of SRLPSO into ELM to compute the optimal weights and bias to make the MSE minimal.

5. Description of Medical Dataset from UCI Repository

The performance of the proposed SRLPSO-ELM method is experimented on five real benchmark classification problems (UCI Machine Learning Repository). The specification of these problems is listed in Table 1. The training, testing, and validation datasets are randomly regenerated at each trial of simulations according to Table 1 for the proposed SRLPSO-ELM algorithms. The five benchmark datasets on which evaluation results are carried out are Wisconsin Breast Cancer, Pima Indians Diabetes, Heart-Statlog, Hepatitis, and Cleveland Heart Disease, which are available from the UCI Machine Learning Repository [36]. Table 1 summarizes the number of features, instances, and classes for each dataset

TABLE 1: Description of datasets.

Dataset	Number of instances	Number of features	Number of classes
Wisconsin Breast Cancer	699	9	2
Pima Indians Diabetes	768	8	2
Heart-Statlog	270	13	2
Hepatitis	155	19	2
Cleveland Heart Disease	296	13	5

used in this study. All this data information is reproduced for the benefit of easy reference from [16].

Wisconsin Breast Cancer [16]. The dataset was collected by Dr. William H. Wolberg (1989–1991) at the University of Wisconsin-Madison Hospitals. It contains 699 instances

TABLE 2: Parameters for the proposed algorithms.

Parameters	ELM	Parameters	SRLPSO
Initial weights and bias between networks	0	Particle size	50
Learning rate	1	Acceleration constants c_1 and c_2	2.0
Number of neurons in input and output layers	Based on datasets considered	w_{min} and w_{max}	0.4 and 1.0
Maximum iteration	500	Maximum iteration	500

characterized by nine features: (1) clump thickness, (2) uniformity of cell size, (3) uniformity of cell shape, (4) marginal adhesion, (5) single epithelial cell size, (6) bare nuclei, (7) bland chromatin, (8) normal nucleoli, and (9) mitoses, which are used to predict benign or malignant growths. In this dataset, 241 (34.5%) instances are malignant and 458 (65.5%) instances are benign.

Pima Indians Diabetes [16]. The dataset is available at the National Institute of Diabetes and Digestive and Kidney Diseases. It contains 768 instances described by eight features used to predict the presence or absence of diabetes. The features are as follows: (1) number of pregnancies, (2) plasma glucose concentration, (3) diastolic blood pressure, (4) tricep skin fold thickness, (5) serum insulin, (6) body mass index, (7) diabetes pedigree function, and (8) age in years.

Heart-Statlog [16]. The dataset is based on data from the Cleveland Clinic Foundation and it contains 270 instances belonging to two classes: the presence or absence of heart disease. It is described by 13 features (age, sex, chest, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic, maximum heart rate, exercise induced angina, old peak, slope, number of major vessels, and thal).

Hepatitis [16]. The dataset is obtained from the Carnegie-Mellon University and it contains 155 instances belonging to two classes: live or die. There are 19 features (age, sex, steroid, antivirals, fatigue, malaise, anorexia, big liver, liver film, palpable spleen, spiders, ascites, varices, bilirubin, alk phosphate, SGOT, albumin, protime, and histology).

Cleveland Heart Disease [16]. The dataset was collected from the Cleveland Clinic Foundation and contains about 296 instances, each having 13 features (originally 76 raw features), which are used to infer the presence (values 1, 2, 3, and 4) or absence (value 0) of heart disease. The features are (1) age, (2) sex, (3) chest pain type, (4) resting blood pressure, (5) cholesterol, (6) fasting blood sugar, (7) resting electrocardiographic results, (8) maximum heart rate, (9) exercise induced angina, (10) depression induced by exercise relative to segment, (11) slope of peak exercise, (12) number of major vessels, and (13) thal.

6. Experimental Results and Discussion

The performance of the proposed SRLPSO-ELM method is experimented on five real benchmark classification problems (UCI Machine Learning Repository). Out of the data samples, 70% is employed for training process and 30% for testing process. The specification of these problems is listed in Table 2.

On applying the proposed SRLPSO-ELM algorithm for the considered medical datasets from the repositories, the dataset features are selected and the parameters, classification accuracy, sensitivity, and specificity, are noted and are tabulated. Hence, in this case, the performance of the proposed method is evaluated in terms of sensitivity, specificity, and accuracy. In general, sensitivity and specificity are the statistical measures employed basically to carry out the performance of classification functions. A perfect classifier or a predictor will be described as 100% sensitive and 100% specific. Since the considered application is of medical dataset involving complex data, the classification should be carried out in an accurate manner. Henceforth, sensitivity and specificity are chosen to be the parametric indices for carrying out the medical dataset classification.

Sensitivity (true positive fraction) is the probability that a diagnostic test is positive, given that the person has the disease:

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{10}$$

Specificity (true negative fraction) is the probability that a diagnostic test is negative, given that the person does not have the disease:

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{11}$$

Accuracy is the probability that a diagnostic test is correctly performed:

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \tag{12}$$

where TP (true positives) is correctly classified positive cases. TN (true negative) is correctly classified negative cases. FP (false positives) is incorrectly classified negative cases. FN (false negative) is incorrectly classified positive cases.

6.1. Breast Cancer Dataset. In Table 3, the results are reported for different feature selection methods for the breast cancer dataset. On classifying the dataset employing original features, it is noted that the classification accuracy of 95.85%, sensitivity of 0.92, and a specificity of 0.98 are obtained. When the PSO-ELM approach is applied an accuracy of 99.62%, sensitivity of 0.9961, and a specificity of 0.9893 are obtained, respectively. On applying the proposed SRLPSO and ELM approach, the accuracy is increased significantly to 99.78%. The best sensitivity and specificity of 1.00 are achieved

TABLE 3: Classification results with Breast Cancer dataset.

| Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
Optimized LVQ (10x CV) [10]	96.70	91.29	92.34	2, 3, 6
Big LVQ (10x CV) [10]	96.80	95.23	96.10	2, 3, 6
AIRS (10x CV) [10]	97.20	96.92	95.00	2, 3, 6, 7
Supervised fuzzy clustering (10x CV) [11]	95.57	98.23	97.36	2, 3, 6, 7, 8
Fuzzy-AIS-knn (10x CV) [12]	99.14	99.56	100	2, 3, 6, 7, 8
F-score + support vector machine [13]	99.51	99.24	98.61	2, 3, 6, 7
Association rule + neural network [14]	97.4	93.12	91.26	2, 3, 6, 7, 8
Artificial metaplasticity neural network [15]	99.26	100	97.89	2, 3, 6, 7, 8
Mean selection method [16]	95.99	93	97	2, 3, 6, 7
Half selection method [16]	96.71	94	98	2, 3, 6, 7, 8
Neural network for threshold selection [16]	97.28	94	99	1, 2, 3, 5, 6, 7, 8
PSO + ELM	**99.62**	**99.61**	**98.93**	**2, 3, 6, 7, 8**
Proposed SRLPSO + ELM	**99.78**	**100**	**100**	**2, 3, 6, 7**
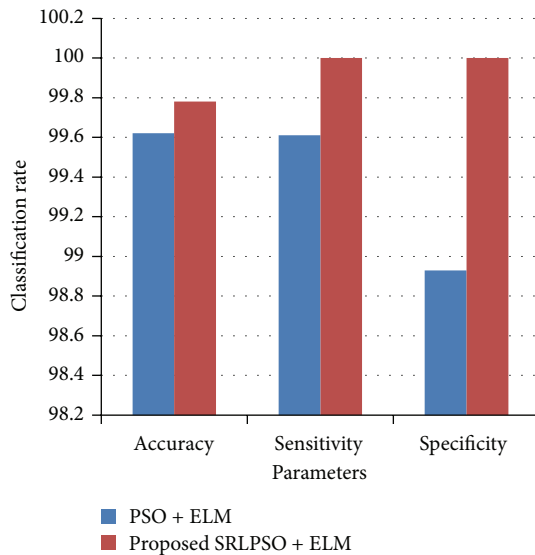
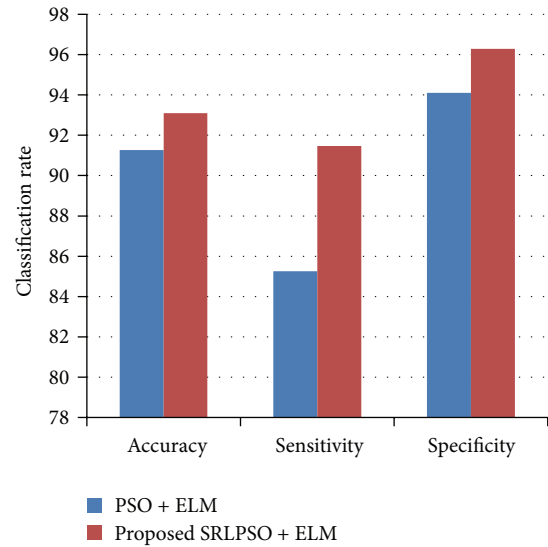FIGURE 4: Classification rate for Breast Cancer dataset.

FIGURE 5: Classification rate for Diabetes dataset.

with the SRLPSO and ELM approach. The highest accuracy is reported for this dataset when the proposed SRLPSO-ELM approach is employed. Figure 4 shows the classification rate for breast cancer dataset employing PSO and proposed SRLPSO with ELM classifier with respect to accuracy, sensitivity, and specificity.

6.2. Pima Indians Diabetes Dataset.

The performance of the different feature selection methods for the Pima Indians Diabetes dataset is shown in Table 4. It is noted that, on applying the basic PSO approach with that of ELM, in comparison with the other methods accuracy and the other parameters are increased to a value of 91.27% for accuracy and 0.8526 and 0.9410 for sensitivity and specificity. Employing the proposed SRLPSO with ELM mechanism has resulted in accuracy significantly increasing to 93.09%. Also in the case of proposed SRLPSO-ELM methodology, the sensitivity and specificity are noted to be 0.9147 and 0.9629, respectively,

with that of only three features, wherein eight features were required in case of the original dataset. Figure 5 shows the classification rate for Diabetes dataset employing PSO and proposed SRLPSO with ELM classifier with respect to accuracy, sensitivity, and specificity.

6.3. Heart-Statlog Dataset.

Table 5 depicts the results obtained by the different feature selection methods with the Heart-Statlog dataset. The classification accuracy using the proposed SRLPSO-ELM has increased to 89.96% with three features. The required features were reduced drastically achieving better classification accuracy. Sensitivity is noted to be 0.8779 and specificity is 0.8842, comparatively better than all the other earlier existing techniques as well as that of PSO-ELM. The proposed SRLPSO-ELM achieves better classification accuracy rate than the earlier methods from the literature considered for comparison. Figure 6 shows the classification rate for Heart-Statlog dataset employing PSO

TABLE 4: Classification results with Pima Indians Diabetes dataset.

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
PCA-ANFIS (10x FC) [17]	89.47	70	71.1	1, 2, 6, 7, 8
LS-ELM (10x FC) [18]	78.21	73.91	80	1, 2, 6, 8
GDA-LS-ELM (10x FC) [18]	79.16	79.1	83.33	1, 2, 6, 7, 8
MLNN with LM (10x FC) [19]	79.62	70	70.31	1, 2, 6, 8
PNN (10x FC) [19]	78.05	71	70.5	2, 6, 8
LDA-MWELM [20]	89.74	83.33	93.75	1, 2, 6, 7, 8
Mean selection method [16]	76.04	71	78	2, 6, 8
Half selection method [16]	75.91	69	79	1, 2, 6, 8
Neural network for threshold selection [16]	76.04	71	78	2, 6, 8
PSO + ELM	**91.27**	**85.26**	**94.10**	**1, 2, 6, 8**
Proposed SRLPSO + ELM	**93.09**	**91.47**	**96.29**	**2, 6, 8**

TABLE 5: Classification results with Heart-Statlog dataset.

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
Evolutionary sigmoidal unit neural network (ESUNN) [21]	83.22	84.32	81.65	3, 8, 9, 11, 12
Evolutionary product unit neural network (EPUNN) [21]	81.89	83.67	84.91	8, 9, 11, 12
Multilogistic regression + EPUNN [22]	83.12	78.15	80.59	8, 9, 11, 12, 13
Mean selection method [16]	84.44	85	84	3, 8, 9, 11, 12, 13
Half selection method [16]	84.81	85	84	3, 8, 9, 10, 11, 12, 13
Neural network for threshold selection [16]	85.19	85	86	3, 11, 12, 13
PSO + ELM	**85.88**	**86.00**	**86.03**	**3, 11, 12, 13**
Proposed SRLPSO + ELM	**89.96**	**87.79**	**88.42**	**11, 12, 13**

FIGURE 6: Classification rate for Heart-Statlog dataset.
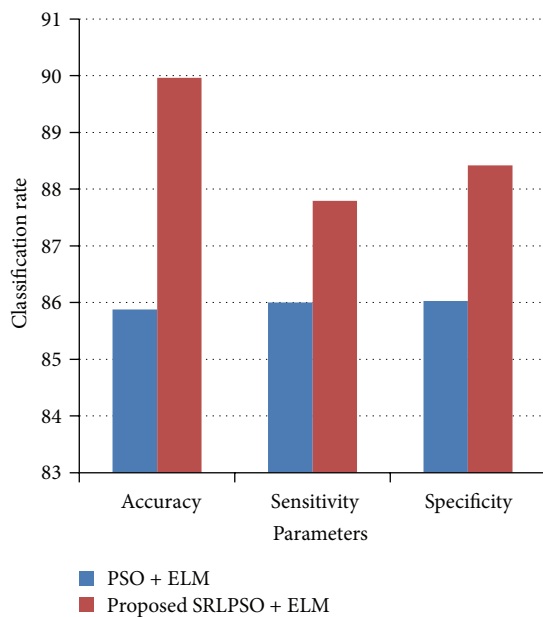

FIGURE 7: Classification rate for Hepatitis dataset.

and proposed SRLPSO with ELM classifier with respect to accuracy, sensitivity, and specificity.

6.4. Hepatitis Dataset. The results for the Hepatitis dataset are shown in Table 6. From the results, it is inferred that the proposed SRLPSO-ELM approach yields a better accuracy of 98.71%, sensitivity 0.9427, and specificity 0.9604. Conventionally, with that of the original features an accuracy of 84.52%, sensitivity of 0.90, and specificity of 0.63 were noted. The proposed SRLPSO-ELM approach achieved the accuracy rate of 98.71% with six features. Figure 7 shows the classification rate for Hepatitis dataset employing PSO and
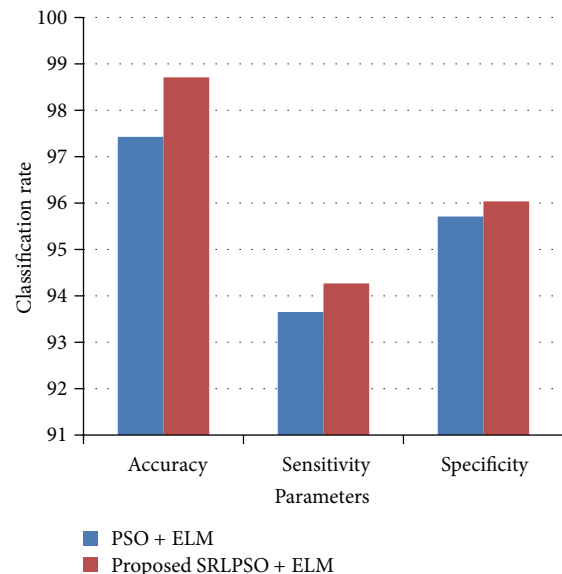
Table 6: Classification results with Hepatitis dataset.

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
Conventional artificial neural network [23]	97.00	92.31	94.5	All
Mean selection method [16]	82.58	87	60	5, 6, 11, 12, 13, 14, 17, 19
Half selection method [16]	85.16	90	66	2, 5, 6, 10, 11, 12, 13, 14, 17, 19
Neural network for threshold selection [16]	85.16	90	66	2, 5, 6, 10, 11, 12, 13, 14, 17, 19
PSO + ELM	**97.43**	**93.65**	**95.71**	**6, 11, 12, 13, 14, 17, 19**
Proposed SRLPSO + ELM	**98.71**	**94.27**	**96.04**	**11, 12, 13, 14, 17, 19**

Table 7: Classification results with Cleveland Heart Disease dataset.

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
C4.5 [31]	81.11	77.23	76.58	All
Naive Bayes [31]	81.48	80.97	81.22	3, 9, 11, 12
BNND [31]	81.11	82.13	80.42	3, 11, 12
BNNF [31]	80.96	76.93	75.81	3, 8, 9, 11, 12
AIRS [24]	84.50	75.34	72.96	All
Hybrid neural network [26]	87.40	93.00	78.50	3, 8, 9, 10, 11, 12
Neural networks ensemble [25]	89.01	80.95	95.91	All
Mean selection method [16]	81.75	82	82	3, 8, 9, 11, 12, 13
Half selection method [16]	83.44	84	83	3, 8, 9, 10, 11, 12, 13
Neural network for threshold selection [16]	84.46	82	82	3, 12, 13
PSO + ELM	**89.47**	**94.49**	**96.02**	**9, 10, 11, 12, 13**
Proposed SRLPSO + ELM	**91.33**	**95.46**	**97.29**	**11, 12, 13**

proposed SRLPSO with ELM classifier with respect to accuracy, sensitivity, and specificity.

6.5. Cleveland Heart Disease Dataset. The results for the Cleveland Heart Disease dataset can be seen in Table 7. As a five-class classification problem is dealt instead of a binary-class classification problem, the Region of Convergence is omitted. It is to be noted that a classification accuracy of 83.82%, sensitivity of 0.83, and a specificity of 0.85 were noted when the original features of the dataset were considered. On applying the proposed SRLPSO-ELM approach, the best accuracy of 91.33% is achieved. This accuracy is achieved with only three features, compared with that of the 13 features of the original dataset. Figure 8 shows the classification rate for Cleveland Heart Disease dataset employing PSO and proposed SRLPSO with ELM classifier with respect to accuracy, sensitivity, and specificity.

7. Conclusion

In this research a new hybrid algorithm that integrates the proposed self-regulated particle swarm optimization (SRPSO) algorithm with the extreme learning machine (ELM) for classification problems is presented. To optimize the input weights and hidden biases and minimum norm least-square scheme, an improved PSO is used to analytically determine the output weights. The PSO is enhanced by incorporating a mechanism which diversifies the search behavior of the particles so that the algorithm finds much better solutions. The performance of the proposed ELM framework
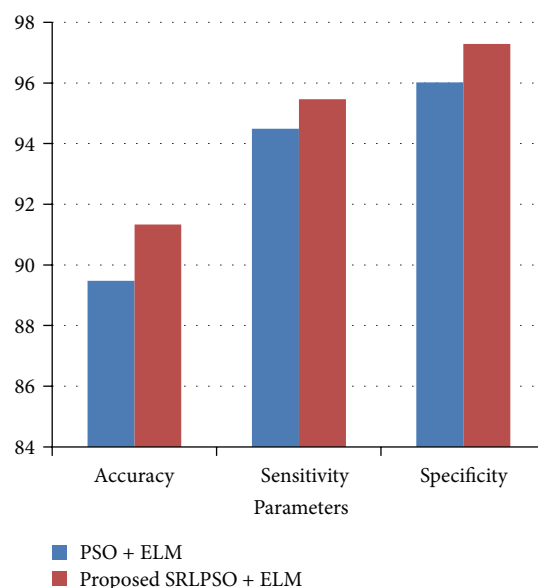
Figure 8: Classification rate for Cleveland Heart Disease dataset.

using SRLPSO was better than the performance of the other methods reported in the literature for five benchmark datasets from the UCI Machine Learning Repository which are used for evaluation. The results also show that in the proposed framework the number of neurons in the hidden layer does not need to be selected by trial-and-error and the relevant input variables can be automatically selected, thus reducing the network size and improving the generalization capability.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] G. S. Babu and S. Suresh, "Meta-cognitive RBF network and its projection based learning algorithm for classification problems," *Applied Soft Computing Journal*, vol. 13, no. 1, pp. 654–666, 2013.

[2] S. N. Sivanandam and S. N. Deepa, *Principles of Soft Computing*, Wiley India Ltd, New Delhi, India, 2007.

[3] G.-B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, 2008.

[4] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[5] T. Matias, F. Souza, R. Araújo, and C. H. Antunes, "Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine," *Neurocomputing*, vol. 129, pp. 428–436, 2014.

[6] D. Chyzhyk, A. Savio, and M. Graña, "Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI," *Neurocomputing*, vol. 128, pp. 73–80, 2014.

[7] Q.-Y. Zhu, A. K. Qin, P. N. Suganthan, and G.-B. Huang, "Evolutionary extreme learning machine," *Pattern Recognition*, vol. 38, no. 10, pp. 1759–1763, 2005.

[8] F. Han, H.-F. Yao, and Q.-H. Ling, "An improved evolutionary extreme learning machine based on particle swarm optimization," *Neurocomputing*, vol. 116, pp. 87–93, 2013.

[9] S. Salcedo-Sanz, C. A. Casanova-Mateo, A. Pastor-Sánchez, and M. Sánchez-Girón, "Daily global solar radiation prediction based on a hybrid coral reefs optimization—extreme learning machine approach," *Solar Energy*, vol. 105, pp. 91–98, 2014.

[10] D. E. Goodman Jr., L. C. Boggess, and A. B. Watkins, "Artificial immune system classification of multiple-class problems," in *Proceedings of the Artificial Neutral Networks in Engineering Conference (ANNIE '02)*, pp. 179–184, November 2002.

[11] J. Abonyi and F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2195–2207, 2003.

[12] S. Şahan, K. Polat, H. Kodaz, and 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, no. 3, pp. 415–423, 2007.

[13] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3240–3247, 2009.

[14] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3465–3469, 2009.

[15] A. Marcano-Cedeño, J. Quintanilla-Domínguez, and D. Andina, "WBCD breast cancer database classification applying artificial metaplasticity neural network," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9573–9579, 2011.

[16] P. Jaganathan and R. Kuppuchamy, "A threshold fuzzy entropy based feature selection for medical database classification," *Computers in Biology and Medicine*, vol. 43, no. 12, pp. 2222–2229, 2013.

[17] K. Polat and S. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Digital Signal Processing*, vol. 17, no. 4, pp. 702–710, 2007.

[18] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 482–487, 2008.

[19] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8610–8615, 2009.

[20] D. Çalişir and E. Doğantekin, "An automatic diabetes diagnosis system based on LDA-wavelet support vector machine classifier," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8311–8315, 2011.

[21] F. J. Martínez-Estudillo, C. Hervás-Martínez, P. A. Gutiérrez, and A. C. Martínez-Estudillo, "Evolutionary product-unit neural networks classifiers," *Neurocomputing*, vol. 72, no. 1–3, pp. 548–561, 2008.

[22] C. Hervás-Martínez, F. J. Martínez-Estudillo, and M. Carbonero-Ruz, "Multilogistic regression by means of evolutionary product-unit neural networks," *Neural Networks*, vol. 21, no. 7, pp. 951–961, 2008.

[23] G. Reibnegger, G. Weiss, G. Werner-Felmayer, G. Judmaier, and H. Wachter, "Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 24, pp. 11426–11430, 1991.

[24] K. Polat, S. Sahan, H. Kodaz, and S. Güneş, "A new classification method to diagnosis heart disease: supervised artificial immune system (AIRS)," in *Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN '05)*, 2005.

[25] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.

[26] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82–89, 2008.

[27] A. Slowik and M. Bialko, "Training of artificial neural networks using differential evolution algorithm," in *Proceedings of the Conference on Human System Interaction (HSI '08)*, pp. 60–65, IEEE, May 2008.

[28] S.-W. Lin, T.-Y. Tseng, S.-Y. Chou, and S.-C. Chen, "A simulated-annealing-based approach for simultaneous parameter optimization and feature selection of back-propagation networks," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1491–1499, 2008.

[29] J. Yu, S. Wang, and L. Xi, "Evolving artificial neural networks using an improved PSO and DPSO," *Neurocomputing*, vol. 71, no. 4–6, pp. 1054–1060, 2008.

[30] S. Salcedo-Sanz, A. Pastor-Sánchez, L. Prieto, A. Blanco-Aguilera, and R. García-Herrera, "Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—extreme learning machine approach," *Energy Conversion and Management*, vol. 87, pp. 10–18, 2014.

[31] N. Cheung, *Machine learning techniques for medical analysis [M.S. thesis]*, University of Queensland, 2001.

[32] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, IEEE, Perth, Australia, December 1995.

[33] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufman Publishers, San Francisco, Calif, USA, 1st edition, 2001.

[34] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proceedings of the IEEE International Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, pp. 69–73, Anchorage, Alaska, USA, May 1998.

[35] K. Kothavari, B. Arunadevi, and S. N. Deepa, "A hybrid DE-RGSO-ELM for brain tumor tissue categorization in 3D magnetic resonance images," *Mathematical Problems in Engineering*, vol. 2014, Article ID 291581, 14 pages, 2014.

[36] A. Frank and A. Asuncion, "UCI Machine Learning Repository," University of California, School of Information and Computer Science, Irvine, Calif, USA, 2010, http://archive.ics.uci.edu/ml.

Advances in
Multimedia

The Scientific
World Journal

International Journal of
Distributed
Sensor Networks

Journal of
Industrial Engineering

Applied
Computational
Intelligence and Soft
Computing

Advances in
Fuzzy
Systems

Modelling &
Simulation
in Engineering

Journal of
Computer Networks
and Communications

Advances in
Artificial
Intelligence

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Advances in
Computer Engineering

International Journal of
Computer Games
Technology

International Journal of
Biomedical Imaging

Advances in
Artificial
Neural Systems

Advances in
Software Engineering

Journal of
Robotics

Advances in
Human-Computer
Interaction

Computational
Intelligence and
Neuroscience

International Journal of
Reconfigurable
Computing

Journal of
Electrical and Computer
Engineering